AD-A048 568 DECISIONS AND DESIGNS INC MCLEAN VA F/G 5/11
EXPERIMENTAL TESTS OF INDEPENCENCE ASSUMPTIONS FOR RISKY MULTIA--ETC(U)
OCT 76 D VON WINTERFELDT N00014-76-C-0074

UNCLASSIFIED NL

1 of 2

ADA048 568

FG.

001813-6-T

social science research institute

UNIVERSITY OF SOUTHERN CALIFORNIA

USC

TECHNICAL REPORT

EXPERIMENTAL TESTS OF INDEPENDENCE ASSUMPTIONS FOR RISKY MULTIATTRIBUTE PREFERENCES

DETLOF VON WINTERFELDT

SPONSORED BY:
ADVANCED RESEARCH PROJECTS AGENCY
DEPARTMENT OF DEFENSE

MONITORED BY:
ENGINEERING PSYCHOLOGY PROGRAMS
OFFICE OF NAVAL RESEARCH
CONTRACT NO. N00014-76-C-0074, ARPA

OCTOBER 1976

SSRI RESEARCH REPORT 76-8

(9) Technical Report 76-8

(6) EXPERIMENTAL TESTS OF INDEPENDENCE ASSUMPTIONS
FOR RISKY MULTIATTRIBUTE PREFERENCES.

by

(10) Detlof von Winterfeldt

(15) N00014-76-C-0074

Sponsored by

Defense Advanced Research Projects Agency
Under Subcontract from
Decisions and Designs, Incorporated

(11) Oct[redacted] [redacted]976

(12) 196 p.

SOCIAL SCIENCE RESEARCH INSTITUTE
University of Southern California
Los Angeles, California 90007

390 664

nt

SUMMARY

The purpose of this experiment was to analyze models of human preferences in complex decision situations that are characterized by uncertainty and multiple attributes of outcomes. Four basic models for such risky multiattribute preferences were considered, among them the additive and multiplicative expected utility models. Independence assumptions that can test the descriptive validity of these models were formulated.

The validity of the independence assumptions, including the marginality assumption and utility independence, was tested for subjects' preferences among even chance gambles for commodity bundles containing gasoline and ground beef. Subjects matched gambles or commodity bundles against a standard and these matches were checked to see if the indifference held in various stimulus contexts as required by the independence assumptions. Effects of response modes, instructions, and personal preference characteristics were examined.

All independence assumptions and models were violated by a bias to prefer a gamble or commodity bundle that was previously matched against a standard, independently of context. Systematic and strong violations of the marginality assumption were found in form of a multivariate risk aversion: subjects tended to prefer a gamble with more balanced multiple outcomes over a gamble with extreme multiple outcomes, even if all single outcomes had an equal chance of occurring. Both the bias and multivariate risk aversion were independent of response modes and instructions. Other preference characteristics such as single attribute risk attitude and preferential interaction of commodities seemed unrelated to multivariate risk aversion.

The bias to prefer a previously matched gamble over a standard cannot be explained by any traditional model describing risky multiattribute preferences. This bias could be due either to mismatching or to a change in preferences after matching. The phenomenon of multivariate risk aversion proved to be a stable property of risky multiattribute preferences for the stimuli considered. Descriptive models for risky multiattribute preferences will have to take this phenomenon into account in similar stimulus situations. For normative modelling the results of the experiment indicate the necessity to carefully check the consistency of preferences assessed by procedures that are based on indifference judgments and to compare

them with actual choices. The multivariate risk aversion effect suggests that simple additive expected utility models may, in some cases, be inappropriate for prescribing preferences. Checks of the marginality assumption and analyses of the form of multivariate risk aversion should be designed and tested carefully, before modelling decision makers' preferences with additive expected utility models.

ACKNOWLEDGEMENTS

I would like to thank the members of my dissertation committee for their support in the preparation of this thesis. I wish to express my gratitude to Ward Edwards for his guidance, help, and encouragement during my studies of decision theory. Special thanks are due to Kanda Kunze who helped me to bring this thesis into a presentable format.

Preparation of this thesis was supported by the Advanced Research Projects Agency, Department of Defense, under the Department of the Navy, ONR, prime contract with Decisions and Designs, Inc., subcontract P.O. 75-030-0711.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

INTRODUCTION

Some of the most interesting and common difficulties in human decision making arise when decisions are risky and their outcomes have several value aspects or attributes. A decision is said to be risky if the decision maker cannot predict with certainty which outcome it will produce. Gambling and buying insurance policies are the most typical examples for decision making under risk. Should one buy collision insurance for an old and rusty VW? Or should one bet on a good looking long shot in a horse race? Such decision situations confront the decision maker with the problem of trading off probabilities of outcomes against their benefits and losses. These trade-offs are inherently subjective, depending on the decision maker's present assets, his risk attitudes, his likelihood estimates and his subjective evaluations of costs and benefits. There are no clear cut rules or formulas, and risky decision making problems can rarely be solved without much difficult deliberation.

A decision problem is called multiattributed if decision outcomes vary in more than one value relevant dimension or attribute. Examples are buying a new car, renting an apartment, allocating funds to social programs, or industrial facility siting. Is an ocean view really worth an additional $50 in rent for an apartment? Or is a 10 mile increase in the distance of an industrial site from a densely populated

area worth a \$100,000 increase in the costs of building access roads? Solving such multiattribute problems requires trade-offs among benefits and losses in one attribute against benefits and losses in another. As in risky decision making such trade-offs are subjective and no clear cut rules exist for making them. How a decision maker trades off outcome attributes against others depends on his subjective judgments of the relative importance of attributes and his evaluations of outcomes within each attribute.

Since risky and multiattribute decision problems are common, problematic, and inherently subjective, they have been the most frequently studied paradigms in behavioral decision theory. Other problems such as time or group preferences have received comparatively little attention. In the late 50's and the early 60's most theoretical and experimental research concentrated on the single attribute risky paradigm, i.e., on decision problems that are characterized by single-dimensional outcomes that are uncertain. Since the end of the 60's, decision theoretic research has shifted to problems of multiattribute decision making. But while most of the theoretical research dealt with the <u>joint</u> risky and multiattribute decision problem in which outcomes are both risky and multiattributed, experimental work was almost exclusively restricted to the riskless multiattribute paradigm in which outcomes are multiattributed but certain. The experiment described in this thesis tried to fill this gap

in the experimental literature. Its purpose was to provide a deeper understanding of the behavioral phenomena that characterize preferences in joint risky and multiattribute decision problems.

Before discussing the theoretical and experimental literature of this joint problem of risky multiattribute preferences, it is useful to summarize briefly the main results on the separate topics of risky single attribute preferences and riskless multiattribute preferences. Most of the experimental literature on human decision making under risk examined subjects' preferences among gambles for money as a prototypical decision problem. The most prominent models to describe such risky preferences are expectation models which assume that subjects order gambles according to some expected value, where subjective transformations may be allowed on probabilities and on monetary values. Expectation models vary in the degree to which they allow such transformations, ranging from the strict expected value model (EV) that does not allow either transformation to the subjective expected utility model (SEU) that allows both. v. Neumann and Morgenstern (1947) laid the foundation of expectation models, that was later modified by Savage (1954) and Luce and Krantz (1971). Alternative models to describe human preferences under risk are portfolio theory (Coombs, 1972) and minimax or minimax regret models (see Wendt, 1970).

In portfolio theory, subjects are assumed to trade-off the expected monetary value of a gamble against some perceived quantity, called the "risk" of the gamble. Minimax and minimax regret models completely ignore probabilities and require that decisions are made on the basis of outcomes alone.

Psychological research has almost exclusively studied expectation models. Most of the experimental results are summarized in Edwards (1954, 1961), Becker and MacClintock (1967), and Lee (1971). The main conclusions of these experiments were:

1) When compared with global preferences among gambles for monetary outcomes, expectation models describe subjects' preferences well.
2) Experimental situations can be created in which subjects consistently violate most of the assumptions and implications of expectation models, including many invariance assumptions (sure thing, irrelevance of context gambles, etc.) and even transitivity.

This contradiction between global validity and specific violations has not been resolved up to now. Proponents of alternative theories (Coombs, 1972) consider expectation models invalid in principle. Other researchers stress the global validity of expectation models which they consider hardly shaken by the specific evidence against them. No matter what position one takes, however, the fact remains

that expectation models have at present no serious and well tested contender among descriptive models for human preferences.

Riskless multiattribute preference models are mainly decomposition models that express how the values of single attribute evaluations are aggregated to a total value of a multiattribute outcome. The single most prominent model of this class is the weighted additive model in which numerical single attribute evaluations are weighted by an attribute importance weight and added across attributes. The model assumes that the order of these weighted sums describes the actual preference order of the decision maker. Theoretical foundations of such decomposition models can be found in conjoint measurement theory (Krantz, Luce, Suppes, and Tversky, 1971) and in decomposition forms of bisymmetric measurement models (Fishburn, 1975).

In the experimental research typically a simple weighted additive model was constructed, value indices were generated for a set of outcomes, and the correlation between these value indices and subjects' global numerical evaluations or preference orders was computed to indicate the degree to which the additive model described subjects' preferences. Several such validation experiments were performed in the early 70's, all of which are best summarized in Fischer (1975). Correlations between the value indices generated by the additive model and the

global evaluations or preference rank orderings were usually in the .80's, often in the .90's, depending on the number of attributes. This result has typically been interpreted as a high descriptive validity of additive models in complex evaluation or preference problems.

Another interpretation of these results is that additive models can predict people's preferences not because they are additive, but rather because additive models can approximate non-additive preferences very well. This argument is best developed in Dawes and Corrigan (1974), who conclude that under two conditions additive models will be good predictors of non-additive preferences: if preference judgments are measured with error, and if preferences are conditionally monotone in each attribute (that is, if more in each attribute is preferred to less, independently of other constant attribute values).

These experimental results on risky single attribute preferences and riskless multiattribute preferences are not sufficient to understand the nature of preferences when outcomes are both uncertain and multiattributed. There are many interesting problems that arise specifically from the interaction between risk and multiple attributes that have not yet been studied experimentally, although the rich theoretical literature on risky multiattribute preference models offers several hypotheses for experimental investigation. The experiment described in this thesis was designed to test some of the main assumptions underlying

such models, and to analyze the behavioral consequences of the interaction between risk and multiple outcomes in decision making.

In the following section the joint risky multiattribute paradigm will be discussed in detail, various model forms for risky multiattribute preferences and their behavioral assumptions will be described, and the few experimental results that bear on an understanding of the risky multiattribute problem will be summarized. This discussion will then lead to the formulations of the hypotheses and experimental questions of this thesis, followed by the experimental methods and results. The implications of the theoretical discussions and the experimental results for an understanding of human decision making under risk and with multiple attributes of outcomes will be discussed. Possible problems in real world applications of risky multiattribute preference models will be considered in the light of the experimental data.

JOINT RISKY AND MULTIATTRIBUTE PREFERENCES

This section will formally present the risky multiattribute decision making paradigm, explain the main model classes that describe or prescribe human preferences in this paradigm, and summarize the results of four experiments that tested some of the models. The terminology used in this section was adapted from Krantz, Luce, Suppes and Tverky's "Foundations of Measurement" (1971). Other important sources that may help to understand the model expositions are Fishburn and Keeney (1974), and Keeney and Raiffa (in press).

Paradigm

To recap: a risky and multiattribute decision problem is a problem in which a decision maker has to select among alternative courses of action that produce uncertain multiattributed outcomes. Many important individual or organizational decision problems are of this type. An individual decision making example is the problem of job selection, where jobs may have many value relevant aspects such as salary, location, staff benefits, professional climate. Some of these aspects may be uncertain, such as the chances for promotion or salary increases. An organizational example is the siting of industrial plants. Sites vary on many strongly weighed cost and benefit attributes such as costs for building transmission lines, costs of

land acquisition, environmental costs, geological qualities, etc. The risks involved in siting decisions (e.g., risk of earthquakes) and the uncertainty of being able to measure all outcomes precisely (e.g., the exact costs) make this multiattribute problem also risky.

A paradigmatic way to structure the problem of selecting among risky and multiattributed alternatives uses vectors and matrices as descriptions of risky multiattribute alternatives. Consider, for example, that the choice alternatives are gambles for commodity bundles containing certain amounts of gasoline and ground beef. A specific gamble may pay off 10 gallons of gasoline <u>and</u> 8 pounds of ground beef, if heads come up on the flip of a coin, or nothing at all, if tails come up. This gamble can be described by a 2 × 2 matrix in which the rows are single attribute gambles, and the columns are single event outcomes:

$$
\begin{array}{c}\text{A}\\ \text{T}\\ \text{T}\\ \text{R}\\ \text{I}\\ \text{B}\\ \text{U}\\ \text{T}\\ \text{E}\end{array}
\qquad
\begin{array}{cc}
 & \text{EVENTS} \\
 & \begin{array}{cc}\text{HEADS} & \text{TAILS}\end{array} \\
\begin{array}{c}\text{GAS}\\ \text{BEEF}\end{array} & \begin{bmatrix} 10 & 0 \\ 8 & 0 \end{bmatrix}
\end{array}
$$

If all outcomes are contingent on the same events, risky and multiattribute choice alternatives can be described as n × m matrices:

$$
\begin{array}{ccc}
 & & \text{EVENTS} \\
 & & \begin{array}{cccccc} E_1 & E_2 & \cdots & E_j & \cdots & E_m \end{array} \\
\text{ATTRIBUTES} & \begin{array}{c} A_1 \\ A_2 \\ \vdots \\ A_i \\ \vdots \\ A_n \end{array} & \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2m} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{im} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nj} & \cdots & x_{nm} \end{bmatrix}
\end{array}
$$

The A_i's are the attributes under consideration, the E_j's are the events that determine which multiattributed outcome the decision maker will receive, and the x_{ij}'s are amounts (or other real valued descriptions) in attribute i of the outcome one receives if event j occurs. The multiattribute risky decision paradigm then reduces to the study of choices among matrices of the above form. The set of matrices that is available to the decision maker at the time he has to make his selection(s) is called the choice set, labelled X, with typical elements x, y, z. The i-th row vector, $\bar{x}_i$, is the gamble within the i-th attribute. The j-th column vector, $\underline{x}_j$, is the multiattribute outcome that the decision maker receives if event j occurs. In the following, matrix elements x_{ij} and y_{ij} will be used to distinguish elements of two choice alternatives that are different. If elements are

constant across choice alternatives, they will be labelled a_{ij} or b_{ij}, and subscripts will be dropped, if elements are equal within a choice alternative.

Models and Independence Assumptions

Preference models are numerical descriptions of a decision maker's preferences. They assign numbers to the elements in the choice set X such that the numerical order coincides with the decision maker's natural preference order among elements in X, and such that some numerical properties reflect other qualitative properties of the decision maker's preference order. All preference models for risky multiattribute preferences that will be described in the following paragraphs are based on the fundamental weak order model. The weak order model assumes that there exists a function u from the choice set X into the real numbers R such that the numerical values that u assigns to elements in X preserve the order of the decision maker's preferences. Formally, this weak order model has the following representation:

Weak Order Model

There exists a real valued function u

$$u : X \rightarrow R$$

such that

$$x \succsim y$$

if and only if

$$u(x) \geq u(y)$$

for all $x, y \in X$,

where "$x \gtrsim y$" means "y is not preferred to x", and "$\geq$" is the usual order relation among the real numbers.

Besides some technical assumptions, the main behavioral condition on the preferences relation $\gtrsim$ is transitivity of preferences, formally stated as:

<u>Transitivity</u>

if $\quad x \gtrsim y$

and $\quad y \gtrsim z$

then $\quad x \gtrsim z$

for all $x, y, z \in X$.

Beyond the weak order model for preferences among risky and multiattribute decisions, scores of models exist that decompose u into an aggregate of functions of subsets of events and/or attributes. The behavioral axioms that distinguish these models are called independence assumptions. Independence assumptions express in terms of the preference relation which manipulations of elements in X leave their preference ordering unaffected. For example, a very weak independence assumption could require that deleting an element from a subset S of X does not change the preference ordering of the remaining elements in S. A much stronger independence assumption could postulate that preferences among elements in X be unaffected by scalar multiplication of row or column vectors.

Two main model classes can be distinguished on the basis of independence assumptions called "row independence" and "column independence." According to row independence, preferences among elements in X that vary only in some rows are independent of the particular values in the remaining (constant) rows. Or to put it another way, preferences are determined solely by the marginal (single attribute) probability distributions. Row independence can be formulated as follows:

<u>Row Independence</u>

$$\begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{a}_i \\ \vdots \\ \bar{x}_n \end{bmatrix} \succsim \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{a}_i \\ \vdots \\ \bar{y}_n \end{bmatrix} \quad \text{if and only if} \quad \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{b}_i \\ \vdots \\ \bar{x}_n \end{bmatrix} \succsim \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{b}_i \\ \vdots \\ \bar{y}_n \end{bmatrix}$$

$$\text{for all } x, y \in X, \bar{a}_i, \bar{b}_i$$

$$i = 1,2,\ldots,n.$$

The fact that this formulation of row independence only uses one constant context row $\bar{a}$ is of no importance, since all cases in which preferences are conditional on more than one constant row are implied.

Symmetrically, column independence requires that preferences among elements in X that vary only in some columns be independent of the particular values in the other constant columns. In other words, column independence assumes that preferences among elements in X are dependent only on those events in which these elements vary, and not on those events in which they have identical outcomes. This assumption is the usual sure thing assumption in expectation models, here applied to multiattributed outcomes. Formally column independence can be stated as follows:

Column Independence

$$(\underline{x}_1,\underline{x}_2,\ldots\ldots,\underline{a}_j,\ldots\ldots,\underline{x}_n) \succsim (\underline{y}_1,\underline{y}_2,\ldots\ldots,\underline{a}_j,\ldots\ldots,\underline{y}_n)$$

if and only if

$$(\underline{x}_1,\underline{x}_2,\ldots\ldots,\underline{b}_j,\ldots\ldots,\underline{x}_n) \succsim (\underline{y}_1,\underline{y}_2,\ldots\ldots,\underline{b}_j,\ldots\ldots,\underline{y}_n)$$

for all x, y ε X, $\underline{a}_j$, $\underline{b}_j$, and

$j = 1,2,\ldots,m.$

Again the fact that the j-th column was singled out as a conditioning column is of no importance for the formulation of column independence, which applied to any combination of conditioning columns.

If neither row nor column independence holds, no simple decomposition of u into single attribute or event functions is possible. If row independence holds, but column independence fails, u can be expressed as an additive composition of single row (single attribute) functions v_i. This model

will be called Model 1, and it can be formally stated as follows:

$$\underline{\text{Model 1}} \qquad u(x) = \sum_{i=1}^{n} v_i(\bar{x}_i).$$

For example, Model 1 would be an accurate description of a decision maker's preferences if he ordered elements in X purely on the basis of their single attribute certainty equivalents. Model 1 has not yet been studied in the theoretical literature, nor have its theoretical implications been tested. However, Model 1 has been analyzed in connection with expectation models that are based on the column independence assumption.

When outcomes are multiattributed column independence is a necessary condition for all models that express utilities of elements in X as an expected value of single outcome utilities. When applied to multiattributed outcomes, all expectation models assume that preferences are column independent. Models that are based on column independence can be formalized by the following additive combination of single event utility functions u_j:

$$\underline{\text{Model 2}} \qquad u(x) = \sum_{j=1}^{m} u_j(\underline{x}_j).$$

The SEU model (see Savage, 1954; Luce and Krantz, 1971) and all other expectation models are special versions of Model 2, in which $u_j(\underline{x}_j) = c_j u(\underline{x}_j)$; $c_j > 0$:

Model 2 (SEU) $$u(x) = \sum_{j=1}^{m} p(E_j)\cdot u(\underline{x}_j).$$

The additional assumption of the SEU model is a very natural condition, since it is unlikely that the shape of the utility functions u_j should depend on the event E_j. Note, however, that a similar strengthening of Model 1 would not be reasonable, since there is little theoretical basis for assuming that single attribute utility functions have identical shapes.

If preferences satisfy both row and column independence they can be described by Model 3, that combines the SEU version of Model 2 with the additive row Model 1:

Model 3 $$u(x) = \sum_{j=1}^{m} p(E_j)\cdot \sum_{i=1}^{n} u_i(x_{ij}).$$

This model is the special case of Model 2 where $u(\underline{x}_j)$ is an additive function of single attribute utility functions $u_i(x_{ij})$.

Model 3 was first formulated by Fishburn (1965) and Pollak (1967). It was subsequently generalized by Fishburn (1970), and by Fishburn and Keeney (1974). But Fishburn arrived at Model 3 not by assuming a combination of row and column independence, but by a slightly different path. His theory postulated that a utility function over elements in X satisfy the SEU model. By adding assumptions that are special versions of row independence, Model 3 could then be proven. Later research by Keeney (1968, 1974), Fishburn

(1973, 1974), and Farquhar (1974) used this route to develop some weaker decomposition forms for $u(\underline{x}_j)$.

To explain this line of thought, consider that column independence is valid in its SEU-version. Also consider the case in which row independence in its general form is violated. There are various weaker assumptions about preferences that lead to decompositions of the expected utility function u in Model 2.

One such assumption is called <u>utility independence</u>, formulated by Pollak (1967) and Keeney (1968, 1974). As a special case of row independence, utility independence is weaker than the row independence assumption formulated above. Utility independence requires row independence to hold only if the conditioning row that is identical across the compared matrices has equal elements in all events. Formally utility independence can be stated as follows:

<u>Utility Independence</u>

$$\begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{a}_i \\ \vdots \\ \bar{x}_n \end{bmatrix} \succsim \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \\ \bar{a}_i \\ \vdots \\ \bar{y}_n \end{bmatrix} \quad \text{if and only if} \quad \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{b}_i \\ \vdots \\ \bar{x}_n \end{bmatrix} \succsim \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{b}_i \\ \vdots \\ \bar{y}_n \end{bmatrix}$$

$$\text{for all } x, y \in X, \bar{a}_i, \bar{b}_i,$$

$$a_{ij} = \text{const}, \; b_{ij} = \text{const}$$

$$i = 1,2,\ldots,n, \qquad j = 1,2,\ldots,m.$$

If the decision maker's preferences satisfy utility independence, they can be described by a multiplicative model:

<u>Model 4</u> $$1 + ku(x) = \sum_{j=1}^{m} p(E_j) \cdot \prod_{i=1}^{n} [1 + ku_i(x_{ij})].$$

If utility independence holds, another weak version of row independence called marginality (Raiffa, 1969), value independence (Fishburn and Keeney, 1974), or additive independence (Keeney and Raiffa, in press) can prove Model 3. Marginality requires row independence to hold only for gambles that have equally likely events E_j. That is when $p(E_j) = 1/m$. Marginality postulates that a decision maker be indifferent among gambles with equally likely events if they can be made identical by exchanging elements x_{ij} within rows. In other words, permuting row elements in x should not change its utility to the decision maker. Formally, marginality can be stated as follows:

<u>Marginality</u>

$$\begin{array}{c} \begin{array}{ccccc} E_1 & \cdots & E_j & \cdots & E_m \end{array} \\ \begin{bmatrix} x_{11} & \cdots & x_{1j} & \cdots & x_{1m} \\ x_{21} & \cdots & x_{2j} & \cdots & x_{2m} \\ \vdots & & \vdots & & \vdots \\ x_{i1} & \cdots & x_{ij} & \cdots & x_{im} \\ \vdots & & \vdots & & \vdots \\ x_{n1} & \cdots & x_{nj} & \cdots & x_{nm} \end{bmatrix} \end{array} \sim \begin{array}{c} \begin{array}{ccccc} E_1 & \cdots & E_j & \cdots & E_m \end{array} \\ \begin{bmatrix} x'_{11} & \cdots & x'_{1j} & \cdots & x'_{1m} \\ x'_{21} & \cdots & x'_{2j} & \cdots & x'_{2m} \\ \vdots & & \vdots & & \vdots \\ x'_{i1} & \cdots & x'_{ij} & \cdots & x'_{im} \\ \vdots & & \vdots & & \vdots \\ x'_{n1} & \cdots & x'_{nj} & \cdots & x'_{nm} \end{bmatrix} \end{array}$$

For all $x \varepsilon X$ with $p(E_i)$ = const, where x'_{ij} is generated by permuting elements x_{ij} within rows.

Although there are many other forms of models that apply to the risky multiattribute decision making situation, Models 1 and 2 are in a sense the most fundamental ones. All other decomposition forms that have been developed theoretically so far are special cases of the SEU Model 2 in which independence assumptions imply certain decompositions of the single outcome utility function $u(\underline{x}_j)$. Of these models the additive and multiplicative Models 3 and 4 are by far the most prominent ones, and have been thoroughly studied in theory and applications. Model 1 itself has not received much theoretical attention, nor has it been applied in many real world decision contexts.

Figure 1 gives an overview of the most important independence assumptions and models for risky multiattribute preferences. Models and independence assumptions are organized in a flow chart, indicating which independence assumptions have to be satisfied for a model to be valid. Diamonds stand for independence assumptions (tests), and rectangles stand for models.

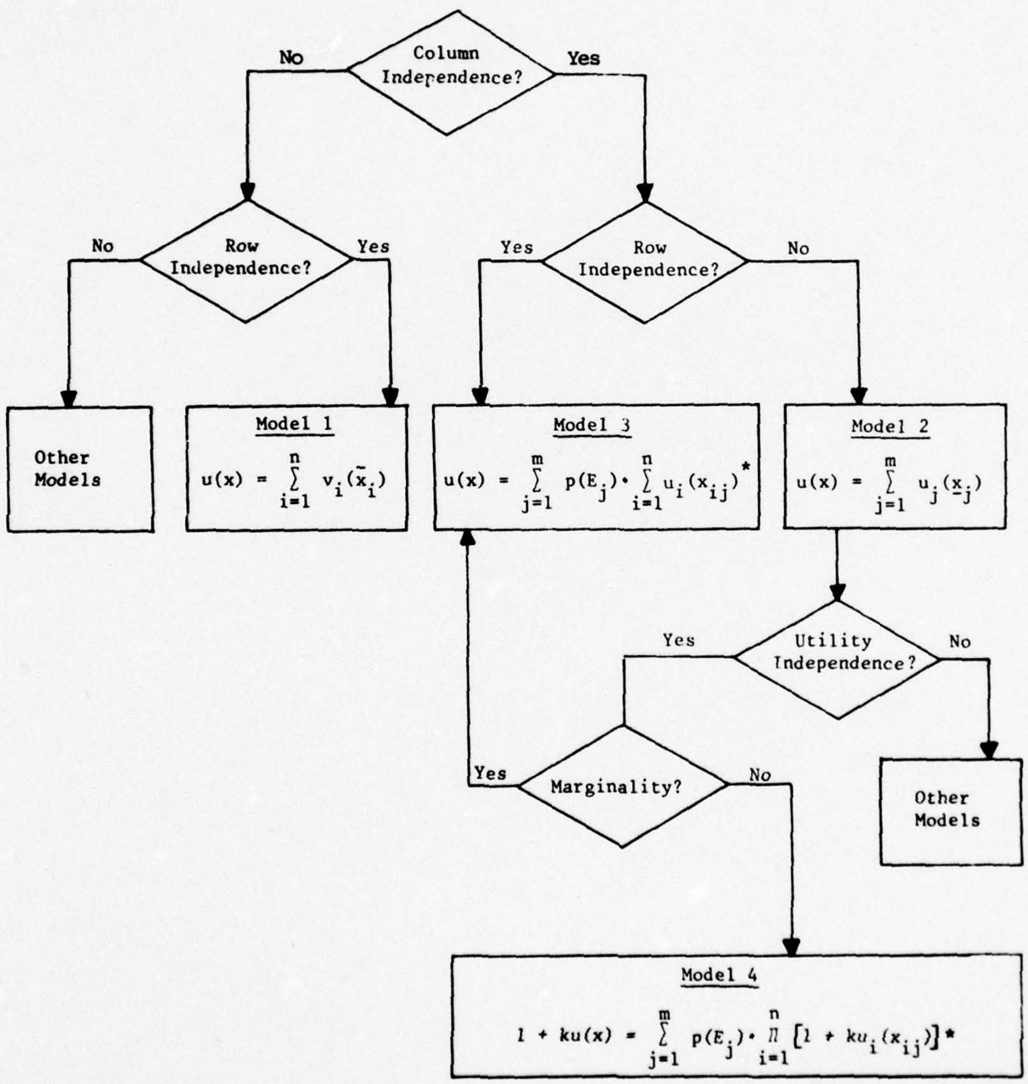
Experiments

Although Models 3 and 4 have been quite extensively applied in decision analyses to solve real world decision problems (see Keeney and Raiffa, in press), very little evidence from controlled experimentation exists on the

Figure 1

Independence Assumptions and Models

for Risky Multiattribute Preferences



*if $u_j(\underline{x}_j) = c_j \cdot u(\underline{x}_j)$ for all j.

ability of decomposition Models 1-4 to describe human preferences. Merely four experiments have any implications for a distinction between these models: Tversky (1967), v. Winterfeldt (1971), Fischer (1972), and Delbeke and Fauville (1974).

Tversky analyzed subjects' preferences for gambles that had two outcomes, cigarettes and candy. Subjects expressed their preferences by stating selling prices for several combinations of cigarettes and candy, as well as for several gambles for either outcome. The results, as far as they bear on a distinction between risky and multiattribute preferences models, are:

1) An additive model could describe the subjects' riskless preferences for combinations of cigarettes and candy very well.

2) If one assumed that riskless utility functions derived from the subjects' preferences among sure commodity bundles were identical up to a linear transformation to risky utility functions derived from subjects' preferences among gambles for either commodity, then subjective probabilities did not sum to unity. If, on the other hand, one assumed that probabilities summed to 1, then one had to acknowledge that risky and riskless utility functions deviated consistently.

3) In spite of these apparent violations of SEU, the overall predictions of an additive SEU model in form of Model 2 fit the bids for gambles with both commodities as outcomes extremely well.

When Tversky conducted his experiment, models such as Models 3 and 4 were not known in their precise axiomatic foundation presented here. In the light of these axioms, Tversky's results could be interpreted differently today. For example, in risky multiattribute preferences there is absolutely no need for risky and riskless utility functions to coincide. Consequently, the difference between risky and riskless utility functions that Tversky detected may not be due to a violation of SEU, but rather to a violation of additivity of risky utility functions in Model 3. If one assumes a non-additive model, as in Model 4, that difference appears quite natural even in the presence of the SEU model. The more recent risky multiattribute models and independence assumptions formulate the real differences between the models more precisely and lend themselves to experimental tests.

v. Winterfeldt (1971) directly analyzed such independence assumptions. When subjects were asked to state their preferences among apartments that varied only on two attributes (distance from campus and transportation facilities), their responses supported a riskless additive multiattribute model. But when asked to state preferences among gambles for such apartments that were designed according to the

marginality assumption, subjects violated the risky additive Model 3 systematically. In spite of such systematic violations of marginality, an additive combination of utility functions u_i in the form of Model 3 predicted subjects' preferences among apartments very well.

The third experiment that studied risky and multiattribute preferences was performed by Fischer (1972). Fischer analyzed subjects' preferences among used cars by determining their global preferences for gambles with used cars as outcomes. He found that the utility function u(x) that was constructed from subjects' risky preferences over gambles with used cars as outcomes deviated significantly from additivity, indicating a violation of Model 3. However, as in v. Winterfeldt's experiment, Model 3 generated a utility function that agreed well with the utility function derived from subjects' global preferences.

Delbeke and Fauville (1974) conducted the most recent experiment. The explicit purpose of their study was to test Fishburn's marginality assumption. Subjects had to state their preferences among marginally equivalent gambles for various two-attribute stimuli, ranging from vacation trips that varied on the attributes location and mode of transportation, to combinations of job types and salaries. Although the small number of subjects and stimuli do not allow conclusive interpretations of their results, Delbeke and Fauville identified some violations of marginality, even if riskless preferences were additive.

The message that these experiments convey seems contradictory: in spite of obvious model violations (tests of marginality and tests of risky additivity failed), additive models such as Model 3 predict subjects' preferences and utility judgments very well. This is the same result that was encountered previously for the simpler riskless multiattribute models and risky single attribute models. The quality in prediction of global preference behavior may similarly be interpreted in terms of the robustness of simple additive models, which describe preferences very well, even if these preferences are not additive and violate model assumptions.

Results of further global model validations seem, therefore, rather predictable, no matter what the type of stimulus or the personal characteristics of the subjects may be. Simple additive expectation models will describe subjects' overall preferences rather well, although they are obviously wrong in some predictions. The more interesting questions that remain are: where are models likely to fail, and for what reason? For example, are Tversky's results due to a genuine failure of the SEU model, or are they a consequence of the non-additive risky utility functions? What is the relative degree of violations of row and column independence in Models 1 and 2? How do these violations depend on the stimulus context? To what degree are violations dependent on instructions and response modes?

Knowing the answers to these questions not only will help in understanding and characterizing human decision making behavior under risk and multiple attributes of outcomes, but may also improve models and assessment methods that are used as aids in real world decision problems. The specific purpose of the present experiment was to test the independence assumptions--row and column independence, utility independence, and marginality--that were described in the previous section. The experiment was designed to establish the degree, the form, and the stability of violations of row and column independence in risky multiattribute preferences, and to determine whether such violations depend on personal preference characteristics of the decision maker.

EXPERIMENTAL ANALYSIS

Experimental Questions and Hypotheses

Four specific questions were singled out for an experimental analysis of the four independence assumptions discussed above:

1) Is there any difference in the degree of violation between row and column independence? The hypothesis was that row independence will be more strongly violated than column independence since subjects may show variance preferences for gambles that have identical marginal probability distributions. Column independence appears to have a firmer foundation in both theory and the experimental results that support expectation models for single and multiattributed outcomes. To test the relative degree of independence violations, symmetric tests for row and column independence with special marginality and utility independence forms were developed for the experiment.

2) What is the form of column or row independence violations, and to what degree do these violations depend on stimulus subsets? In particular, the question arises how the violations of independence assumptions are related to the context rows or columns. Understanding the effect of these

context rows or columns is a key to understanding the characteristics or risky multiattribute preferences. To establish the form of independence violations, stimulus subsets were selected to magnify possible context effects.

3) How stable are violations of row and column independence? Possible violations are likely to depend on response modes and instructions. In Tversky's experiment, for example, the selling price response mode may have forced subjects to think in terms of exchange values and respond in terms of additive cash equivalents, when asked for selling prices for two commodities. To test the dependence of independence violations on response modes, three different response modes were selected for the experimental tests: preferences, ratings, and cash equivalent assessments (selling prices).

4) Are violations of row or column independence related to other personal preference characteristics such as risk attitude, marginal utilities in the single attributes, or additivity/non-additivity of riskless perferences? To investigate the relationship between such personal preference characteristics and independence violations, various indices of risk attitude were assessed and an index for additivity was determined.

Method

Stimuli. One of the major problems of any experiment that tries to learn about the structure of human preferences is the selection of appropriate stimuli. Preferences often depend very much on the physical properties of stimuli, often more than on the personal preference characteristics of the subject/decision maker. A good example are preferences among seats in a football stadium. There exists a physical measure of "visibility" that is probably a complex, but non-additive function of row and gate numbers. Preferences for most visitors will differ little from that physical measure. If one wants to learn about the subjective nature of preferences, seats in a football stadium would therefore not be very interesting, since preferences among seats could be modelled without using any subjective inputs. All one may learn from observing preferences among seats in a football stadium is how well a subject can estimate that physical measure of visibility.

On the other extreme, some stimuli have so little physical structure that preferences are very idiosyncratic. An example is preferences among combinations of coffee, sugar, and milk. Few subjects will agree on a "best" mix, and few will exhibit similar combination rules. Such stimuli may, in fact, be interesting to study experimentally, but there is little hope to ever generalize across subjects in statements about model validity.

In principle, any study of human preferences is subjective and can make statements only about individual results. But there are differences in the degree of generalization depending on the nature of the stimuli. In the football stadium case, one could probably generalize from a pool of subjects to the preferences of many (which one could have without experimentation); in the coffee case, it would be very difficult to say anything beyond individual preference structures. In the design of a study of subjective preference properties like the ones this experiment studies, one would like to be able to make intersubject statements about possible degrees and forms of model violations, while still examining a basically subjective process. The selection of stimuli determines to which degree such balance is possible.

One way to achieve the right balance is by studying many different stimuli and to try to abstract those aspects of preferences that are independent of the actual stimulus, or that are functions of interactions between the stimuli and preferences. Unfortunately, at the present stage of knowledge about risky and multiattribute preferences, such an approach would be very uneconomical, since the experimental questions have not been phrased precisely enough. An alternative is to select a prototypical stimulus with some well-known physical characteristics that still leave room for the manifestation of subjective preferences. Since

experimental results should be generated by the subjective nature of preferences, the physical nature of such stimulus should be amendable to various forms of risky multiattribute models to avoid preclusion of experimental results.

For the present experiment, this latter approach was chosen mainly for economical reasons. The stimuli that were finally selected for the experiment are even chance gambles for market baskets containing certain amounts of gasoline and ground beef. A pilot study revealed for most subjects strong deviations from an additive cash equivalent model (in which subjects simply add the market price of both commodities), marginally decreasing utilities for beef, and somewhat less marginally decreasing utilities for gasoline. All of these properties are compatible with Models 1-4.

Stimuli were presented to subjects in a matrix format. Figure 2 gives an example. The choice set X in this experiment existed, therefore, of 2 × 2 matrices with cell elements representing certain amounts of gasoline and ground beef.

To test row and column independence, certain special stimulus forms had to be created. The general idea of performing these tests was to first establish an indifference between two gambles in one constant row or column context, and then to check if that indifference held up under a different row or column context. For example, to test row

Figure 2

Example Stimuli

Take a careful look at the following two gambles. Assume that you have the right to play one of these gambles once.

G A M B L E 1

if the outcome of a coin flip is

	HEADS		TAILS
	you win		
GALLONS OF GASOLINE	8	or	16
	and		and
POUNDS OF GROUND BEEF	10	or	0

G A M B L E 2

if the outcome of a coin flip is

HEADS		TAILS	
you win			
0	or	16	GALLONS OF GASOLINE
and		and	
10	or	10	POUNDS OF GROUND BEEF

independence, two row stimuli would be matched by the subject, by varying the amount x in one row:

$$\begin{array}{lcccc} & \multicolumn{2}{c}{\text{Gamble 1}} & & \multicolumn{2}{c}{\text{Gamble 2}} \\ & \text{Heads} & \text{Tails} & & \text{Heads} & \text{Tails} \\ \text{Gasoline} & (10 & 0) & \sim & (x & x) \end{array}$$

Then a row independent subject should be indifferent between the following two check stimuli:

$$\begin{array}{lcccc} & \text{Gamble 1} & & \text{Gamble 2} \\ & \begin{array}{cc}\text{Heads} & \text{Tails}\end{array} & & \begin{array}{cc}\text{Heads} & \text{Tails}\end{array} \\ \begin{array}{l}\text{Gas} \\ \text{Beef}\end{array} & \begin{bmatrix} 10 & 0 \\ a & b \end{bmatrix} & \sim & \begin{bmatrix} x & x \\ a & b \end{bmatrix} \end{array}$$

no matter what the constant amounts a and b in the context row are.

Similarly, to test column independence, first two columns are matched by the subject, e.g.,

$$\begin{array}{lccc} & \text{Bundle 1} & & \text{Bundle 2} \\ \begin{array}{l}\text{Gas} \\ \text{Beef}\end{array} & \begin{pmatrix} 10 \\ 0 \end{pmatrix} & \sim & \begin{pmatrix} 0 \\ x \end{pmatrix} \end{array}$$

Then a column independent subject should be indifferent between the following two check stimuli:

$$\begin{array}{lccc} & \text{Gamble 1} & & \text{Gamble 2} \\ & \begin{array}{cc}\text{Heads} & \text{Tails}\end{array} & & \begin{array}{cc}\text{Heads} & \text{Tails}\end{array} \\ \begin{array}{l}\text{Gas} \\ \text{Beef}\end{array} & \begin{bmatrix} 10 & a \\ 0 & b \end{bmatrix} & \sim & \begin{bmatrix} 0 & a \\ x & b \end{bmatrix} \end{array}$$

independently of the amounts a and b in the context column.

To construct check stimuli in this "match and check" procedure, two similar matches were selected for row and column tests, one match labelled EE (for extreme-extreme) and one labelled EM (for extreme-middle). The EE matches involve a substantial stimulus change or trade-off, while the EM matches require less trade-off. Table 1 lists these four matches. One additional EM match was created to futher test row independence in its utility independence form.

Table 1

List of Matches for the Match and Check Procedure

	Row Matches	Column Matches
Extreme-extreme	Gas $\begin{array}{cccc} H & T & & H \quad T \\ (16 & 0) & \sim & (0 \quad 16) \end{array}$	$\begin{array}{l} \text{Gas} \\ \text{Beef} \end{array} \begin{array}{c} \text{sure} \\ \begin{pmatrix} 0 \\ 10 \end{pmatrix} \end{array} \sim \begin{array}{c} \text{sure} \\ \begin{pmatrix} x \\ 0 \end{pmatrix} \end{array}$
Extreme-middle	Gas $\begin{array}{cccc} H & T & & H \quad T \\ (16 & 0) & \sim & (x \quad x) \end{array}$	$\begin{array}{l} \text{Gas} \\ \text{Beef} \end{array} \begin{array}{c} \text{sure} \\ \begin{pmatrix} 0 \\ 10 \end{pmatrix} \end{array} \sim \begin{array}{c} \text{sure} \\ \begin{pmatrix} x \\ 5 \end{pmatrix} \end{array}$
Extreme-middle	Beef $\begin{array}{cccc} H & T & & H \quad T \\ (10 & 0) & \sim & (x \quad x) \end{array}$	

It was not possible to construct row and column matches in perfect symmetry. One reason was that the EE match for rows follows trivially from the equal likelihood of the events Heads and Tails, while the EE match for columns needs to be assessed by the subject. Another reason was that the

EM match in rows has equal amounts, while the EM match for columns did not necessarily have that property. Nevertheless, the matches were created as symmetrically as possible.

The five matches in Table 1 were coupled with nine possible context rows or columns that were generated from combinations of the following single commodity values:

Gasoline	:	0, 8, 16 Gallons
Ground Beef	:	0, 5, 10 Pounds.

Forty-five check items that included pairs of gambles that should be indifferent according to either row or column independence were created this way. The following examples show how these check items can test row independence, utility independence, marginality, as well as similar forms of column independence.

The EE row match coupled with its context rows tests the marginality assumption nine times:

$$\begin{array}{ccc} & \text{Gamble 1} & & \text{Gamble 2} \\ & \begin{array}{cc}\text{Heads} & \text{Tails}\end{array} & & \begin{array}{cc}\text{Heads} & \text{Tails}\end{array} \\ \begin{array}{l}\text{Gas} \\ \text{Beef}\end{array} & \begin{bmatrix} 16 & 0 \\ a & b \end{bmatrix} & \sim & \begin{bmatrix} 0 & 16 \\ a & b \end{bmatrix} \end{array}$$

for a, b = 0, 5, 10.

The EM row matches coupled with their context rows test row independence in its pure form and utility independence. In the gasoline match, for example, check items are of the following type:

$$\begin{array}{cccc} & \text{Gamble 1} & & \text{Gamble 2} \\ & \begin{array}{cc}\text{Heads} & \text{Tails}\end{array} & & \begin{array}{cc}\text{Heads} & \text{Tails}\end{array} \\ \begin{array}{l}\text{Gas} \\ \text{Beef}\end{array} & \begin{bmatrix}16 & 0 \\ a & b\end{bmatrix} & \sim & \begin{bmatrix}x & x \\ a & b\end{bmatrix}\end{array}$$

for a, b = 0, 5, 10.

If a = b, these check items test utility independence. The remaining check items test row independence neither in its marginality nor in its utility independence form. Similar tests are generated for the additional EM row matches in the beef commodity:

$$\begin{array}{cccc} & \text{Gamble 1} & & \text{Gamble 2} \\ & \begin{array}{cc}\text{Heads} & \text{Tails}\end{array} & & \begin{array}{cc}\text{Heads} & \text{Tails}\end{array} \\ \begin{array}{l}\text{Gas} \\ \text{Beef}\end{array} & \begin{bmatrix}a & b \\ 10 & 0\end{bmatrix} & \sim & \begin{bmatrix}a & b \\ x & x\end{bmatrix}\end{array}$$

The column check items are very similar in form to the row check items. The EE matches create column check items that correspond to the marginality test:

$$\begin{array}{cccc} & \text{Gamble 1} & & \text{Gamble 2} \\ & \begin{array}{cc}\text{Heads} & \text{Tails}\end{array} & & \begin{array}{cc}\text{Heads} & \text{Tails}\end{array} \\ \begin{array}{l}\text{Gas} \\ \text{Beef}\end{array} & \begin{bmatrix}0 & a \\ 10 & b\end{bmatrix} & \sim & \begin{bmatrix}x & a \\ 0 & b\end{bmatrix}\end{array}$$

for a = 0, 8, 16; and b = 0, 5, 10.

The other nine check column items are:

$$
\begin{array}{lcc}
 & \text{Gamble 1} & & \text{Gamble 2} \\
 & \begin{array}{cc}\text{Heads} & \text{Tails}\end{array} & & \begin{array}{cc}\text{Heads} & \text{Tails}\end{array} \\
\begin{array}{l}\text{Gas} \\ \text{Beef}\end{array} & \begin{bmatrix} 0 & a \\ 10 & b \end{bmatrix} & \sim & \begin{bmatrix} x & a \\ 5 & b \end{bmatrix}
\end{array}
$$

for $a = 0, 8, 16$; and $b = 0, 5, 10$.

Subjects. Subjects were 18 upper level students (mostly graduates) and staff members from the University of Southern California. All subjects could use gasoline and ground beef for consumption. They were paid a flat fee of $12 for a total of 3-5 hours of participation in the experiment. In addition, they had a chance to win up to 16 gallons of gasoline, 10 pounds of ground beef, or $20 in gambles that were actually played after the experiment in accordance with their preferences.

Instructions and Response Modes. Subjects were presented with the 45 check stimuli twice. In the first run they did not receive any special instructions. In the second run six subjects repeated the first run, six were cued to compare the gambles in a check stimulus by comparing their rows first, and six were cued to compare their columns first. To force subjects to attend to rows (columns) when comparing gambles in check items, the rows or columns that constituted the pair of gambles were presented to them on a page to the left of the check item page and they were asked to mark their preferences among these rows (columns) first, and then to

look at the check item. They were told to carefully consider their preferences among the rows (columns) when making their preference judgments about the combined gambles.

To quantify the subjects' preferences, and to determine if response modes influenced the violations of independence assumptions, three response modes were used: simple preferences or indifferences, a rating of the strength of preferences, and assessments of cash equivalents for both gambles in each check item. In the first response mode subjects simply had to mark the preferred gamble in each check item. In the second response mode they marked the strength of their preference for the preferred gamble on a continuous scale that varied from 0 (indifference) to 100 (very strong preference). In the final response mode subjects were asked to independently assess for each gamble of a check stimulus an amount of money such that they would be indifferent between playing the gamble or receiving the sure amount of money as a gift. To reinforce subjects to state their true indifference point Marschak's bidding procedure was applied and it was explained to subjects that some of the gambles would be played after the experiment according to Marschak's procedure (see Becker, DeGroot, and Marschak, 1964). Figure 3 gives an example of a response sheet to one check item.

Procedure. Subjects were run individually in two sessions. The first session lasted approximately 2-3 hours. The second session lasted between 1½ to 2 hours. At the

Figure 3

Example of a Response Sheet

Take a careful look at the following two gambles. Assume that you have the right to play one of these gambles once.

GAMBLE 1

If the outcome of a coin flip is

	HEADS		TAILS
	you win		
GALLONS OF GASOLINE	16	or	0
	and		and
POUNDS OF GROUND BEEF	0	or	10

GAMBLE 2

If the outcome of a coin flip is

HEADS		TAILS	
you win			
16	or	0	GALLONS OF GASOLINE
and		and	
10	or	0	POUNDS OF GROUND BEEF

1. MARK THE GAMBLE THAT YOU WOULD PREFER TO PLAY IN THE SPACE BELOW. (If you are indifferent, mark both spaces.)

   ____________ ____________

2. INDICATE ON THE FOLLOWING SCALE HOW MUCH YOU PREFER THAT GAMBLE RELATIVE TO THE OTHER ONE. (0 means "no preference"; 100 means "very strong preference".)

   0 20 40 60 80 100

3. STATE YOUR CASH EQUIVALENTS FOR BOTH GAMBLES IN THE SPACES BELOW. (Remember to check if you are really indifferent between playing the gamble or receiving your cash equivalent as a gift.)

   $________ $________

beginning of the experiment subjects received a detailed 15 page instruction booklet, which contained a brief introduction to the experiment, examples of the types of questions that would be asked, and a description of the experimental procedure. Subjects were instructed that their task was to express their preferences among a series of pairs of gambles and to make indifference judgments about them. They were told that some of these gambles would be played after the experiment according to their judgments and preferences, and that they could win up to 16 gallons of gasoline, 10 pounds of ground beef, or $20 if they expressed their true preferences. Subjects were warned that simplifying strategies could hurt their prospective winnings if they did not coincide with their true preferences. Besides these general instructions, several examples for stimuli and response modes were presented and explained in the booklet. After the subjects read the instructions, the experimenter made sure that they understood them, and again warned them to express their true preferences. Marschak's bidding procedure was explained verbally once more in two examples to make sure that all subjects understood its meaning.

Subsequently subjects were presented with training booklet I, which contained several example problems, assessments of certainty equivalents for gas and beef gambles, and assessments of various cash equivalents for complex gambles. All cash equivalents for the bundles including the amounts

0, 5, and 10 pounds of beef and 0, 8, and 16 gallons of gasoline were assessed. Also all five matches required in the experiment were determined. All indifference judgments were motivated by Marschak's bidding procedure.

After completing training booklet I, subjects were asked to describe their strategies to arrive at a preference judgment and to determine cash equivalents for commodity bundles and gambles. Strategies were probed by the experimenter in some examples. Stress was put on expression of true preferences. Certain consistency requirements were pointed out such as agreement between preferences and cash equivalents.

The next task was to respond to training booklet II which was identical to the first booklet, except for some examples. Subjects were told that some of the questions were identical but that they should feel free to change their answers if the discussion had changed their mind. The assessment of certainty equivalents for cash gambles, of cash equivalents for commodity bundles, and of all matches were repeated in this booklet.

After completing the second training booklet, subjects were told that five judgments (matches) that they had previously made were particularly important in the experiment. These five matches were then re-assessed interactively with the experimenter. The experimenter selected a value for the match x and asked the subject to state a preference among the pair. Changing values of x the experimenter slowly focussed

in on a range in which subjects switched their preference. This fact was pointed out to subjects, and they were asked to set their indifference point. At each stage the experimenter encouraged subjects to revise their previous answers, if they felt that they did not reflect their thinking any more.

The first session was concluded with the presentation of three booklets containing the first 45 check items. The order of check items and the position of the gambles in a check item were randomized.

The second session followed within two days after the first. Subjects were presented with the same 45 check items in the same order as in the first session. Six subjects received no special instructions, six received row instructions, and six received column instructions.

Results

Data Description and Outline of Analyses. Each of the 18 subjects was presented with a total of 90 check items (5 matches, 9 contexts, and 2 sessions). For each check item the raw data consisted of:

1) Preference judgment;
2) Rating of the strength of preference;
3) Cash equivalents for the left and the right gamble in the check item (CL and CR).

The following derived measures were used in the analysis:

4) Dollar difference (DD): DD = CL - CR;

5) Absolute dollar difference (ADD): ADD = abs(DD):

6) Percentage dollar difference (PDD):

PDD = DD·100/max(CL,CR).

In addition, the cash equivalents for riskless combinations of gasoline and ground beef assessed in training sessions I and II were used as background information about the subjects' substitution rates between gasoline, beef, and cash.

These data were analyzed on increasing levels of detail. The first analysis was concerned with the consistency of responses. The second analysis compared the subjects' responses with some simple strategies and models. Then the overall effects of response modes, instructions, and items were analyzed on a gross level to sort out the main sources of variation. On the next level a fine analysis of the context dependence of violations of independence assumptions was performed, and finally the relationships between individual parameters characterizing risk attitude, substitution rates, etc. and independence violations was studied.

Consistency Analysis. Since each check item was presented twice, a preliminary consistency analysis could be performed on the changes of preferences that occurred between the two sessions. The first column of Table 2 lists for each subject the percentage of true preference reversals, that is of switches of preferences from the left to the right gamble in a check item or vice versa. Not included are changes from or to indifference. Considering the complexity of the task,

Table 2

Intrasubject Consistency Data of Responses Between the First and Second Experimental Session (n = 36)

Subject	Preference Reversals (%) (n = 45)	Median Absolute Differences of			
		Cash Equivalents	Ratings	Dollar Differences	Percentage Dollar Differences
1	6.7	0.72	8.7	0.34	7.2
2	4.4	0.10	11.5	0.49	9.5
3	15.6	0.09	8.1	0.50	8.6
4	13.3	0.94	10.0	0.20	7.2
5	11.1	0.47	8.8	0.52	10.9
6	0.0	0.06	0.0	0.00	0.0
7	20.0	0.53	2.9	0.60	9.0
8	6.7	0.71	10.0	0.88	13.6
9	4.4	0.95	10.0	0.60	6.9
10	4.4	0.73	18.0	0.45	8.2
11	8.9	0.72	7.2	0.84	15.5
12	2.2	0.71	12.1	0.12	6.9
13	4.4	0.44	7.8	0.27	6.2
14	2.2	0.42	8.2	0.54	7.8
15	6.7	0.43	6.9	0.42	6.2
16	6.7	1.86	12.3	1.98	14.0
17	22.2	0.62	40.0	0.85	9.5
18	4.4	0.47	9.5	0.43	10.0
Median	6.7	0.58	9.2	0.50	8.4

the consistency seems appreciable with a median of 6.7% preference reversals. Eight subjects had less than 5% preference reversals. Only two subjects showed pronounced inconsistencies (subjects 7 and 17). Instruction groups (S1-S6, S7-S12, S13-S18) do not show any marked differences in consistency. Accumulated across subjects the consistency statistics are as follows: 78.4% identical preferences, 13.6% changes from or two indifference, and 8.0% true preference reversals.

The remaining columns of Table 2 present the intrasubject consistency of cash equivalent and rating responses, and of the derived measures, dollar difference and percentage dollar difference. As a consistency measure the median absolute difference of these measures between session one and two was computed. Again, the consistency seems considerable. Only three subjects have a median that is higher than .75 (cash equivalent), 12 (rating), .85 (dollar difference), or 11% (percentage dollar difference). Noteworthy exemptions are subject 16 who appears inconsistent in all measures except rating, and subject 17 whose ratings are very inconsistent.

It is quite possible that the relatively high degree of consistency reflected in Table 2 was a result of simplifying strategies such as always preferring the gamble with the higher sum of cell entries or giving cash equivalents equal to the expected cash value of the gambles. This possibility will be examined next.

Strategy analysis. To check the possibility of simplifying strategies the following simple expected value model was constructed to predict cash equivalent assessments:

$$\widehat{CL} = 1/2(x + y + v + w)$$

where x, y, v, and w are the cell entries of a gamble, and $\widehat{CL}$ is the predicted cash equivalent. Any subject who determined his cash equivalent by some simple arithmetic would produce responses that are highly correlated with CL. Table 3 gives the correlations between CL and $\widehat{CL}$ for each subject. In general the predictive validity of this simple model is high (Median: .70), reflecting once more the stability of simple additive models to predict evaluations that are monotonically related to input variables. However, for most subjects the model validities lie below the subjects reliabilities as computed by the retest correlation between cash equivalent responses in the first and the second session (Median: .79).

There are other indications that subjects deviated from simple expected cash equivalent models. Such models assume neutral risk attitude (expected value assumption), linear substitution of commodities and of commodities with cash, and additivity. To determine if subjects met these assumptions, the following indices were computed:

Table 3

Validities of a Simple Expected Value Model (v) and Reliabilities of Cash Equivalent Assessments (r)

Subject	v (n = 72)	r (n = 36)
1	.80	.85
2	.66	.93
3	.74	.88
4	.61	.68
5	.80	.88
6	.93	.98
7	.78	.74
8	.70	.59
9	.30	.54

Subject	v (n = 72)	r (n = 36)
10	.67	.78
11	.73	.64
12	.55	.79
13	.65	.86
14	.80	.91
15	.70	.78
16	.36	.26
17	.78	.74
18	.65	.87

Median v: 0.70

Median r: 0.79

1) Risk premium for the gasoline gamble (16 0):

$$RP_G = EV(16 \quad 0) - CE(16 \quad 0),$$

where the left term stands for the expected value of the gamble (EV), and the right term for its assessed cash equivalent (CE).

2) Risk premium for the beef gamble (10 0):

$$RP_B = EV(10 \quad 0) - CE(10 \quad 0).$$

If the risk premium is 0 the subject is called risk neutral, if it is larger than 0 he is called risk averse, if it is smaller than 0 risk seeking (see Raiffa, 1968).

3) An indicator for the way *gasoline* and cash substitute:

$$SI_G = \$(16)/\$(8),$$

where $(a) stands for the cash equivalent the subject assessed for <u>a</u> gallons of gasoline.

4) An indicator for the way beef and cash substitute:

$$SI_B = \$(10)/\$(5),$$

where $(b) stands for the cash equivalent the subject assessed for <u>b</u> pounds of ground beef.

In case of linear substitution these ratios would equal 2, they would be smaller than 2 if the commodity is substituted

with money at a decreasing rate, and greater than 2 if it is substituted at an increasing rate.

5) An additivity indicator:

$$AI = \$(a) + \$(b) - \$\begin{pmatrix} a \\ b \end{pmatrix}$$

where $\$\begin{pmatrix} a \\ b \end{pmatrix}$ is the cash equivalent assessed to the commodity bundle containing a gallons of gasoline and b pounds of ground beef; \$(a) and \$(b) are the respective cash equivalents for the single commodities.

This additivity index would be zero if there was no interaction between the commodities, positive if the combination of commodities had a negative effect, and negative if the combination of the commodities had a positive effect on their value. The additivity index used in the further analysis was actually an average of eight indicators of the above form computed from assessments in the first two training sessions. All other indicators were computed from the unique assessments made in the final training session.

Table 4 shows how subjects distribute over these indicators. Even with a rather liberal definition of neutral risk attitude (the risk premium had to fall between -1.00 and +1.00) only four subjects were risk neutral for the gasoline gamble and five for the beef gamble. The remaining subjects distribute relatively evenly into the risk seeking and the risk averse categories. An interesting result was

that six subjects had different risk attitudes for gas and beef according to this classification. One subject was strongly risk averse in the gasoline gamble (RP_G = -.200) but risk seeking in the beef gamble (RP_B = +2.00). Only one subject was perfectly risk neutral in both commodities (RP = 0).

Table 4

Distribution of Subjects According to Risk Attitude, Substitution Rates, and Non-Additivity

	Risk Neutral $-1.00 < RP < 1.00$	Risk Averse $RP \geq 1.00$	Risk Seeking $RP \leq -1.00$
Gas	4	8	6
Beef	5	6	7

	Linear Substitution $1.75 < SI < 2.25$	Decreasing Substitution $SI \leq 1.75$	Increasing Substitution $SI \geq 2.25$
Gas	12	3	3
Beef	5	12	1

Additive $-1.00 < AI < 1.00$	Negative Interaction $AI < -1.00$	Positive Interaction $AI > 1.00$
8	8	2

The distribution over the substitution index (gas with cash) shows that the majority of subjects had close to linear substitution rates. They were willing to exchange the first

8 gallons of gasoline at approximately the same cash amount as an additional 8 gallons. In the beef commodity, however, the majority of subjects showed a decreasing rate of substituting beef for cash. Only one subject had an increasing rate. This means that most subjects would pay more for the first 5 pound package of beef than for the second.

The last box in Table 4 shows that most subjects were close to additive in their cash evaluation of commodity bundles (8 subjects) or displayed a negative interaction (8 subjects).

In summary: Subjects showed pronounced deviations from the expected value model and linear substitution of beef with cash. Only a single subject falls into the categories "risk neutral", "linear substitution", and "additive" in all cases of Table 4 (subject 6).

Overall Effects. The response mode effect was rather limited. Ratings of the strength of preference among gambles were correlated highly enough with dollar differences (Median r: .74) to conclude that they represented similar judgments about the relative value of the gambles. Only one subject had essentially a zero correlation between ratings and dollar differences. She was indifferent between all column check items, but she assessed different cash equivalents for most column pairs. Only two other subjects had correlations between ratings and

dollar differences that were smaller than .66. The preferred gamble had a lower cash equivalent in merely 1.9% of all check items. Therefore, in the following analyses of the effects of instructions, matches, and contexts, only preferences, dollar differences, and absolute dollars differences are used as dependent variables.

The initial question was: what was the overall trend in preferences among check items, and what was the general level of model violations? Already on the level of ordinal preferences a strong effect appeared that was somewhat unexpected. Subjects tended to prefer the gamble that had as one element a matched value. This effect held for most subjects independently of context rows or columns. Table 5 shows the distribution of preferences for each match accumulated across subjects, sessions, and context. Preferences were evenly distributed around indifference only in the case of the 'natural' row match $(16 \;\; 0) \sim (0 \;\; 16)$. In the case of the two other row matches, preferences were heavily biased in the direction of the certainty equivalent side. In the case of the column matches, preferences were biassed toward the matched gasoline value. Only one subject deviated markedly from this general pattern. He always preferred the left gamble of a column check item.

Table 6 tells the same story in terms of mean dollar differences. The bias towards the gamble with the

Table 5

Distribution of Preferences (in %) as a Function of Matches (Across Subjects Sessions, and Contexts; n = 324)

Match	Preferred Gamble: Left	No	Right	Total
$(16\quad 0) \sim (0\quad 16)$	30.2	43.8	25.9	100
$(16\quad 0) \sim (x\quad x)$	11.1	21.6	67.3	100
$(10\quad 0) \sim (x\quad x)$	16.0	16.7	67.3	100
$\begin{pmatrix} 0 \\ 10 \end{pmatrix} \sim \begin{pmatrix} x \\ 0 \end{pmatrix}$	11.7	37.0	51.2	100
$\begin{pmatrix} 0 \\ 10 \end{pmatrix} \sim \begin{pmatrix} x \\ 5 \end{pmatrix}$	16.7	23.5	59.9	100
Total	17.2	28.5	54.3	100

Table 6

Mean Dollar Differences and Standard Errors as a Function of Matches (Across Subjects, Sessions, and Contexts; n = 324) (A minus sign means preference for the right gamble)

Match	Mean	Standard Error
$(16\quad 0) \sim (0\quad 16)$	0.01	0.08
$(16\quad 0) \sim (x\quad x)$	-0.78	0.09
$(10\quad 0) \sim (x\quad x)$	-0.91	0.08
$\begin{pmatrix} 0 \\ 10 \end{pmatrix} \sim \begin{pmatrix} x \\ 0 \end{pmatrix}$	-0.38	0.12
$\begin{pmatrix} 0 \\ 10 \end{pmatrix} \sim \begin{pmatrix} x \\ 5 \end{pmatrix}$	-0.39	0.08

matched amount shows up stronger here for the row matches than for the column matches. Thus, even without considering specific contexts or instructions, a strong preference bias could be shown. Note, however, that these data say nothing about the degree of violations of row and column independence, since low averages may be based on very large dollar differences with opposite signs.

To analyse the degree of row and column independence and its dependence on instructions, preferences and absolute dollar differences were used as dependent variables. Subjects S1-S6 received no instructions in the second session, subjects S7-S12 received row instructions, and subjects S13-S18 received column instructions. The consistency data of Table 2 did not give any indication that the consistency of preference responses was influenced by this grouping. One would expect that under row instructions previous preferences among row check items would become indifferent, and that under column instructions previous preferences among column check items would become indifferent. In fact, no such trend could be discovered. Without instructions 2.5% of all row check item responses (aggregated across subjects and contexts) switched from preference to indifference, under row instructions 4.9%, and under column instruction 9.9%. Without instructions 13% of the column check items responses switched from preference to indifference, under

row instructions 7.4%, and under column instructions 8.3%.

For a more detailed analysis of the effects of row and column items and of instructions, absolute dollar difference was used as a dependent variable. Table 7 presents the individual means (in brackets: standard errors) of ADD for row and column check items in both sessions. Several trends become obvious when inspecting this table. First, the general level of independence violations is not very high (Grand mean across subjects, sessions, and items: $1.08; standard deviation: 1.42). Second, there are notable intersubject differences in the degree and the relative size of the violations of row and column independence. Finally, the variance of ADD for row and column check items seemed to differ frequently, with a tendency of the row variance to be larger than the column variance (exceptions are subjects 14 and 16). An F-test showed that row and column variances differed significantly for 9 subjects ($p < .05$).

Because of the intersubject variability and the different variances of row and column items the session effect was tested separately for each subject and for row and column responses with a t-test for paired samples. Of the 36 t-tests 5 showed a significant difference between mean ADDs for column items, two for row items ($p < .05$). No systematic pattern of instruction effects could be identified in these differences. In the group

Table 7

Individual Means of Absolute Dollar Differences for Row and Column Check Items in Each Session (In brackets: Standard Errors; n=18)

Subject		Session 1	Session 2	$\bar{x}$	t	p
1	Row	1.06 (.17)	1.02 (.14)	1.04 (.11)	4.32	.00
	Col	0.44 (.14)	0.33 (.16)	0.39 (.10)		
2	Row	1.83 (.34)	2.00 (.34)	1.91 (.24)	4.25	.00
	Col	0.81 (.14)	0.88 (.11)	0.84 (.09)		
3	Row	1.28 (.26)	0.94 (.17)	1.11 (.16)	0.50	.61
	Col	1.11 (.22)	0.89 (.23)	1.00 (.10)		
4	Row	1.33 (.29)	0.82 (.20)	1.08 (.17)	4.62	.00
	Col	0.33 (.05)	0.25 (.05)	0.29 (.02)		
5	Row	0.75 (.19)	0.61 (.13)	0.68 (.12)	1.60	.10
	Col	0.61 (.14)	0.28 (.06)	0.45 (.08)		
6	Row	0.00 (.00)	0.00 (.00)	0.00 (.00)	0.00	.50
	Col	0.00 (.00)	0.00 (.00)	0.00 (.00)		
7	Row	1.39 (.32)	1.28 (.28)	1.34 (.21)	1.10	.27
	Col	1.33 (.29)	0.75 (.09)	1.05 (.16)		
8	Row	1.73 (.44)	1.69 (.44)	1.71 (.31)	-0.03	.97
	Col	2.13 (.33)	1.31 (.28)	1.72 (.23)		
9	Row	1.36 (.27)	0.75 (.16)	1.05 (.16)	1.17	.30
	Col	1.05 (.19)	0.64 (.07)	0.85 (.06)		
10	Row	0.75 (.16)	0.63 (.16)	0.69 (.09)	0.14	.92
	Col	0.89 (.18)	0.51 (.12)	0.67 (.11)		
11	Row	0.67 (.17)	1.11 (.26)	0.89 (.16)	1.24	.22
	Col	0.56 (.16)	0.75 (.15)	0.65 (.11)		

Table 7 (continued)

Subject		Session 1	Session 2	$\bar{x}$	t	p
12	Row	0.78 (.17)	1.06 (.25)	0.92 (.15)	-0.92	.47
	Col	1.06 (.06)	1.00 (.00)	1.03 (.03)		
13	Row	1.40 (.29)	1.29 (.23)	1.35 (.18)	7.12	.00
	Col	0.09 (.06)	0.00 (.00)	0.05 (.03)		
14	Row	0.89 (.16)	0.47 (.12)	0.68 (.11)	-12.02	.00
	Col	2.75 (.20)	2.35 (.20)	2.55 (.11)		
15	Row	0.79 (.23)	0.96 (.20)	0.88 (.15)	4.19	.00
	Col	0.11 (.04)	0.37 (.04)	0.23 (.04)		
16	Row	2.78 (.50)	2.50 (.47)	2.64 (.34)	-3.69	.00
	Col	4.83 (.79)	5.67 (.93)	5.25 (.62)		
17	Row	1.21 (.17)	1.87 (.27)	1.54 (.17)	1.63	.14
	Col	1.04 (.23)	1.39 (.16)	1.21 (.11)		
18	Row	0.95 (.26)	0.53 (.26)	0.74 (.18)	2.07	.05
	Col	0.40 (.12)	0.26 (.13)	0.34 (.07)		

Total row : 1.13 (.05)

Total col : 1.03 (.06)

of subjects that did not receive any special instructions, one subject significantly decreased his column violations in the second session. In the row instruction group two subjects decreased their column violations significantly, one decreased his row violations significantly. In the column instruction group one subject increased his column violations, one subject decreased his column violations, and one subject increased his row violations.

To test the difference between mean ADDs of row and column check items, ADD was averaged across sessions. The average ADD for row and column check item are listed in the fourth column of Table 7 together with their standard errors. The last two columns give the t values and the significance level. For 13 subjects row independence was more strongly violated than column independence. Six of these differences are significant at the .05 level or lower. Only for two subjects (S14 and S16) column violations were significantly higher than row violations. Notable is the lack of any violations for S6. Across subjects the difference between row and column violations is insignificant.

In summary: the analysis of overall effects found no substantial response mode effects, strong intersubject variability in the degree and relative size of row and column independence violations, no systematic instruction effects, and only a modest effect of row vs. column

violations that differed among subjects. To understand more deeply the nature of row and column violations, it seems therefore necessary to look into the effects of specific context rows or columns on responses, and to establish the specific pattern of the row and column violations as a function of these contexts. This analysis will be done next.

<u>Context analysis</u>. Tables 8-12 present the most interesting results of the experiment. They show the preference distributions and mean dollar differences (together with their standard errors) for each of the five matches as a function of context row or column. The data are accumulated across subjects and sessions.

First to the 'natural' row match (16 0) $\sim$ (0 16). As one would expect subjects were predominantly indifferent in all those check items that were generated by constant context rows (the first three contexts in Table 8). Only two responses were not indifferent (for the (10 10) context) presumably because the subjects made an error.

The block of the next three context rows represents those check items in which the left gamble had more balanced outcomes than the right. That is in the left gamble the subjects had a chance of winning some modest amount in either event, while in the right gamble they had a chance at winning much vs. nothing. A large marjority of the subjects preferred the left gamble, following

Table 8

Preferences and Mean Dollar Differences

as a Function of Context Rows

(Match: (16 0) ∿ (0 16); n = 36)

Context Row	% Preferences Left	No	Right	Mean DD*	Standard Error
(0 0)	0	100	0	0.00	0.00
(5 5)	0	100	0	0.00	0.00
(10 10)	0	94	6	-0.06	0.06
(0 5)	83	14	3	0.75	0.16
(5 10)	59	22	19	0.90	0.23
(0 10)	81	8	11	1.23	0.31
(5 0)	10	18	72	-0.57	0.15
(10 5)	25	22	53	-0.83	0.34
(10 0)	3	19	78	-1.36	0.22

*A negative sign means preference for the right gamble in the check item.

A general trend towards "multivariate risk aversion", as Richard (1975) called it. This effect is least pronounced for the (5 10) context row, and strong for the (0 5) and (0 10) rows. The mean dollar differences reflect that trend in dollar terms. A positive dollar difference here means a higher cash equivalent for the left gamble. The last three rows in Table 8 show the results for the inverted context rows, which created identical check items but in reversed order. The results reflect that symmetry.

To determine to which degree the pattern of violations in Table 8 reflects individual responses, the number of preferences that agreed with the multivariate risk aversion pattern was counted for each subject. Multivariate risk aversion would predict that a subject responds with indifference in the first three row contexts, with a preference for the left gamble in the second three row contexts, and with a preference for the right gamble in the last three row contexts. Three subjects fitted this pattern perfectly in all of their 18 responses. Eight subjects had only one or two deviations from this pattern, five subjects had three or four deviations, and only two subjects (including the indifferent subject 6) showed more than four violations.

A more refined pattern analysis was based on the individual dollar differences. For each subject the average dollar difference was computed for each of the three groups of context rows in Table 8. If a subject's

cash equivalence responses would agree with the general multivariate risk aversion pattern, his ranking of these three mean dollar differences would be as follows: lowest (negative) for the last three contexts, medium (close to zero) for the first three contexts, and largest (positive) for the middle three. For 15 subjects the mean dollar differences were ranked in exactly this way. Two subjects had reversed rankings, and subject 6 had zero dollar differences for all three context groups.

Tables 9 and 10 present the results for the row matches $(16 \quad 0) \sim (x \quad x)$ and $(10 \quad 0) \sim (x \quad x)$. The pattern of preferences is not quite as clear here, since the bias for the gamble with the matched value x starts to dominate the picture. But while the bias is clearly the main feature for the first and last three context rows, the middle three context rows counteract the bias. For these context rows the left gamble in the check item becomes again more balanced. Correspondingly subjects tended to increase their preferences for this gamble.

To determine how well individual response patterns agreed with the general patterns reflected in Tables 9 and 10, individual average dollar difference were computed for each of the three context row groups in the two tables. If there was a multivariate risk aversion effect counteracting the bias effect, one would expect that these mean dollar differences were ranked in the following order:

Table 9

Preferences and Mean Dollar Differences

as a Function of Context Rows

(Match: (16 0) ∿ (x x); n = 36)

Context Row	% Preferences Left	No	Right	Mean DD	Standard Error
(0 0)	8	12	80	-0.83	0.16
(5 5)	3	22	75	-1.47	0.22
(10 10)	3	16	81	-1.53	0.24
(0 5)	25	14	61	-0.16	0.31
(5 10)	17	19	64	-0.46	0.22
(0 10)	36	14	60	0.32	0.31
(5 0)	8	11	81	-0.54	0.26
(10 5)	3	19	78	-1.89	0.25
(10 0)	11	17	72	-0.43	0.21

Table 10

Preferences and Mean Dollar Differences

as a Function of Context Rows

(Match: (10 0) $\sim$ (x x); n = 36)

Context Row	% Preferences Left	No	Right	Mean DD	Standard Error
(0 0)	6	19	75	-0.53	0.15
(8 8)	11	17	72	-1.16	0.32
(16 16)	12	19	69	-1.76	0.35
(0 8)	22	14	64	-0.25	0.17
(8 16)	14	17	69	-0.82	0.25
(0 16)	34	19	47	-0.12	0.36
(8 0)	11	8	81	-0.70	0.16
(16 8)	6	19	75	-2.00	0.48
(16 0)	6	17	77	-0.87	0.18

lowest for the last three context rows, middle for the first three context rows, and highest for the middle three context rows. This ranking would be expected, since the last three context rows make the left gamble more unbalanced than the first three rows, while the middle three context rows make the left gamble more balanced. For the (16 0) ∿ (x x) match mean dollar differences were ranked this way for four subjects, 11 subjects showed only one reversal, one subject had two reversals, and one had three. For the (10 0) ∿ (x x) match mean dollar differences were ranked according to the multivariate risk aversion pattern for seven subjects, eight subjects had one reversal, and two subjects had two. Subject 6 was again indifferent in all responses and consequently had zero dollar differences in all cases.

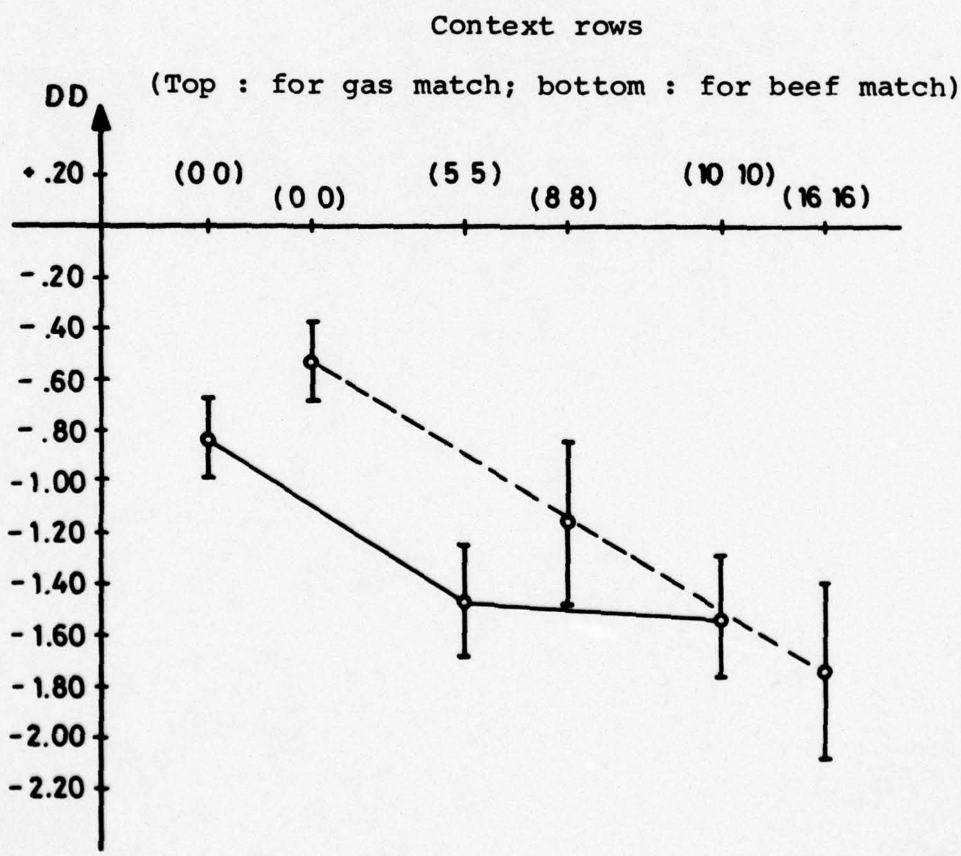
Although the trend for multivariate risk aversion is by no means as pronounced for these last two matches as it was for the natural row match (16 0) ∿ (0 16), it seems to be stable and persistent within subjects. The data then suggest that row indpendence in its pure form is also violated in a systematic and context dependent way.

What about violations of row independence in its utility independence form? Here the first three context rows of Tables 9 and 10 are relevant. Although the preference distributions show no clear trend, the mean

Figure 4

Mean Dollar Differences as a Function
of Context Row for Matches
(16 0) (x x) and (10 0) (x x) (n = 36)

(The lines mark off one standard error)



dollar difference increases as the constant row context increase. Figure 4 presents this trend graphically for both matches. Only the maximum difference for the beef trend is significant.

In the check items that tested column independence the picture is dominated by the bias to prefer the gamble with the matched gasoline amount (see Tables 11 and 12). For neither match a clear context effect is observable. To determine how well individual response patterns agreed with the general pattern reflected in Tables 11 and 12, individual average dollar difference were computed for each of the three context column groups for each subject. If there was any effect of the grouping of context columns, similar to the multivariate risk aversion effect, one would expect some characteristic ranking of these mean dollar differences. In fact, no such typical ranking could be found. Subjects distributed evenly over all possible rankings in both matches. Thus it seems that violations of column independence are largely due to the preference bias and not context dependent.

Individual Parameters and Independence Violations. These last few result paragraphs will discuss some individual characteristics as related to violations

Table 11

Preferences and Mean Dollar Differences

as a Function of Context Columns

(Match: $\binom{0}{10} \sim \binom{x}{0}$; n = 36)

Context Column	% Preferences Left	No	Right	Mean DD	Standard Error
$\binom{0}{0}$	14	33	53	-0.75	0.36
$\binom{8}{5}$	14	28	58	-0.38	0.35
$\binom{16}{10}$	14	36	50	-0.02	0.33
$\binom{0}{5}$	16	31	53	-0.47	0.33
$\binom{0}{10}$	11	36	53	-0.65	0.50
$\binom{8}{10}$	17	39	44	0.14	0.39
$\binom{8}{0}$	11	42	47	-0.39	0.33
$\binom{16}{0}$	19	34	47	-0.41	0.33
$\binom{16}{5}$	14	39	47	-0.47	0.43

Table 12

Preferences and Mean Dollar Differences

as a Function of Context Columns

(Match: $\binom{0}{10} \sim \binom{x}{5}$; n = 36)

Context Column	% Preferences Left	No	Right	Mean DD	Standard Error
$\binom{0}{0}$	11	25	64	-0.29	0.17
$\binom{8}{5}$	19	22	59	-0.49	0.32
$\binom{16}{10}$	17	33	50	-0.41	0.22
$\binom{0}{5}$	14	28	58	-0.30	0.20
$\binom{0}{10}$	17	27	56	-0.35	0.17
$\binom{8}{10}$	14	22	64	-0.55	0.25
$\binom{8}{0}$	8	31	61	-0.34	0.17
$\binom{16}{0}$	25	14	61	-0.50	0.27
$\binom{16}{5}$	22	25	53	-0.29	0.26

of independence assumptions. A correlation analysis related the previously described individual parameters (risk premium for gasoline and beef gambles, substitution indices for gasoline and beef with cash, and additivity indices) with indices of independence violation. These indices are the individual averages in absolute dollar differences for row and column check items (see Table 7 that lists these indices) and an index of 'multivariate risk aversion'. This index was constructed as follows:

$$\text{MRA} = \$\begin{bmatrix} 16 & 0 \\ 0 & 10 \end{bmatrix} - \$\begin{bmatrix} 16 & 0 \\ 10 & 0 \end{bmatrix}$$

where $\$[\quad]$ stands for the assessed cash equivalent. For each subject the index was an average of assessments of four presentations.

A positive MRA index means that subjects tended to prefer the more balanced gamble on the left, a negative index means that they preferred the more imbalanced gamble on the right, and a zero index means indifference. Table 13 presents the correlations of the 5 independent indices with the individual mean absolute dollar differences for row and column items and the MRA index. Since the number of subjects was small, the results have to be interpreted with caution.

Table 13

Correlations Between Indices of Preference Characteristics and Indices of Independence Violations (n = 18)

	Index for		
	Multivariate Risk Attitude	Row Independence Violation	Column Independence Violation
Risk Premium (Gas)	0.41	-0.30	-0.30
Risk Premium (Beef)	-0.20	-0.60	-0.55
Substitution Index (Gas)	0.55	0.42	0.27
Substitution Index (Beef)	0.13	0.37	0.26
Additivity Index	-0.03	0.02	-0.05

The additivity index is uncorrelated with all three indices of independence violations. From non-additivity of commodities one can therefore not draw any conclusions about possible violations of row or column independence. The risk premium of the gasoline gamble (16 0) is correlated positively with the MRA index. This indicates that subjects who are risk averse in a single commodity are also somewhat multivariate risk averse. However, the risk premium of the beef gamble (10 0) is slightly negatively correlated with the MRA index, casting doubt on the

generality of the conclusion that single attribute risk aversion and multivariate risk aversion go hand in hand.

Both risk premiums were negatively correlated with the mean absolute dollar differences of row and column check items. Subjects who were more risk seeking tended to violate independence assumptions more. Furthermore, all substitution indices are positively correlated with the indices of independence violations. Subjects who had a more pronounced decreasing substitution rate between the commodities and money tended to violate row and column independence less.

Much of the previous analysis has dealt with general trends across subjects. In the last few pages, some subjects will be singled out who deviated from such general trends, and their responses will be discussed in more detail. Subject 6 showed a very clear response pattern. He was indifferent in <u>all</u> 90 check items. He was also the one subject who was risk neutral in both commodities, had almost linear substitution rates for gas and beef with cash, and was close to additive in his cash evaluations of commodity bundles. Naturally he had the highest validity of the simple EV model described on page 45. It is clear that subject 6 used a simple pricing and expected value strategy to determine his cash equivalents. As he pointed out in the interview following the experiment, he had recognized the construction principle of the check items and had responded with indifference statements. It was his opinion

that his responses reflected his true preferences, since the stakes were not large enough. When he was presented in a later session with check items in which all amounts were multiplied by 10, he showed exactly the response pattern that most subjects displayed in the actual experiment. In particular, he became risk averse, showed a quite strong decreasing substitution rate of beef with cash, and became multivariate risk averse.

Subject 16 on the other hand had the highest violations of row and column independence. The most outstanding characteristic of his responses was their extremeness. His cash equivalents for commodity bundles were much higher than the cash equivalents of other subjects, overshooting even the market price substantially. He was very strongly risk seeking, and he even gave a certainty equivalent of 12 to the beef gamble (10 0), a response that violated the expected utility principle. He insisted on these extreme responses, however, with the explanation that he liked to gamble and that he did not care for the actual outcomes that much. Besides the extremeness of his responses, subject 16 was also the most inconsistent subject. His retest reliability for cash equivalents was only .26. Thus the high violation of row and column independence for this subject may be explained by his erratic and extreme responses.

Subject 14 showed the second highest violations of column independence. He was the only subject who often

that his responses reflected his true preferences, since the stakes were not large enough. When he was presented in a later session with check items in which all amounts were multiplied by 10, he showed exactly the response pattern that most subjects displayed in the actual experiment. In particular, he became risk averse, showed a quite strong decreasing substitution rate of beef with cash, and became multivariate risk averse.

Subject 16 on the other hand had the highest violations of row and column independence. The most outstanding characteristic of his responses was their extremeness. His cash equivalents for commodity bundles were much higher than the cash equivalents of other subjects, overshooting even the market price substantially. He was very strongly risk seeking, and he even gave a certainty equivalent of 12 to the beef gamble (10 0), a response that violated the expected utility principle. He insisted on these extreme responses, however, with the explanation that he liked to gamble and that he did not care for the actual outcomes that much. Besides the extremeness of his responses, subject 16 was also the most inconsistent subject. His retest reliability for cash equivalents was only .26. Thus the high violation of row and column independence for this subject may be explained by his erratic and extreme responses.

Subject 14 showed the second highest violations of column independence. He was the only subject who often

responded with indifference to a check item, but gave different cash equivalents. In fact, he was indifferent among all but two column check items, thus fitting column independence very well in terms of preferences. But his strategy to determine independently his cash equivalents for both gambles in the check item did not reflect this indifference. The relatively high validity of the simple EV model (.80) indicates that he used some pricing strategy to set cash equivalents.

Subject 13 showed a very low degree of column independence violation. He also had linear substitution rates of commodities with money and was nearly additive in cash equivalent assessments for commodity bundles. His low degree of column violation may have been a consequence of this simple linear additive strategy in commodity bundle evaluations. His risk attitudes were not very strong either. But unlike subject 6 who showed a similar pattern of individual parameters, his multivariate risk aversion was very strong, leading to his violations of row independence.

All other subjects had response patterns that corresponded roughly to the general trends outlined in the previous sections.

DISCUSSION

Before making an attempt to explain the results of the experiment and to discuss their implications for descriptive and normative modelling of risky multiattribute preferences, the question of the interpretability and generalizability of the experimental data will be addressed. Data from choice experiments are susceptible to two kinds of problems: internal response inconsistencies and simplifying response strategies. If subjects respond very inconsistently, the experimental results simply become messy and possible experimental effects are buried under the response errors. If subjects use simplifying strategies, they may be very consistent, but their responses may not reflect their true preferences. In either case the interpretation of the results is seriously hampered.

Fortunately, the experimental results show that neither randomness nor simplifying strategies dominated the response patterns. Except for one subject, all subjects responded with a high degree of internal consistency. This consistency was probably due to the simple construction principle of the check items. Although gambles of the form used in the experiment may be difficult to compare _in general_, the specific gambles that constituted the check items had some prominent features that allowed an easier comparison on the basis of straightforward preference characteristics.

For example, subjects could rather easily decide among pairs of gambles that tested the marginality assumption, if they were multivariate risk averse, and as long as that preference characteristic did not change, their responses would not change in repeated comparisons. As subjects mentioned in the post-experimental interviews, they found the preference tasks simple, mainly because of the ease with which they could establish some basic preference characteristics in their choices. However, most subjects thought that the cash equivalent assessments for gambles was difficult. Why then the relatively high consistency of cash equivalent assessments?

It is possible that subjects used some simple arithmetic to generate cash equivalents. But in the discussion sessions in the training period it became clear that, except for subject 6, all subjects thought that averaging, pricing, or expected value strategies would misrepresent their true preferences. Most subjects described their cash equivalent assessment strategies as an 'averaging and discounting' strategy. In assessing certainty equivalents they would first think about the expected value, realize a preference, and adjust the cash equivalent upward or downward according to the direction of their initial preference. In assessing cash equivalents for commodity bundles they would first use some market price equivalent and then discount for the marginal utility of the commodities. The results reflect

these strategies quite well, in particular in the risk attitudes of most subjects and in the non-linear substitution rates of commodities with cash.

The manifestation of clear preference characteristics such as risk aversion (or risk seeking) and non-linear substitution rates not only leads to the argument that subjects probably did not use simple response strategies. It also says something about the generalizability of the results beyond the simple stimulus that was used in the experiment. If subjects had used strategies that would only be applicable to stimuli that allow linear trade-offs among commodities, and between commodities and cash, and if the outcomes had been so inconsequential as to suggest expected value computation, the generalizability would be very limited. But since subjects did in fact exhibit quite strong preference characteristics, there is reason to assume that the results generalize to such stimuli that create similar preference characteristics, that is all those for which risk attitude, non-linear substitution rates and non-additivity are of a similar form as for the gambles for beef and gasoline.

A first look at the general pattern of the responses to check items revealed a strong bias to prefer the gamble with a matched value. This bias was independent of context, only to some extend dependent on matches, and held for all subjects similarly. One possible explanation is that subjects

mismatched, that is the matched values were different from their true indifference point. An explanation of such mismatching would be an 'anchoring and adjustment' process. Tversky and Kahnemann (1975) describe such processes in judgments about uncertainty, and they hypothesize that several biasses in such judgments can be explained by insufficient adjustment. In the present experiment subjects may have focussed or anchored on the expected value of a gamble or some price equivalent of a commodity bundle, and adjusted that value up or down according to their risk attitude and marginal utilities of commodities. In fact, many subjects described their matching process in somewhat similar terms.

Insufficient adjustment in such an 'anchoring and adjustment' process can explain some of the biasses that were observed in the experiment. Since subjects had to match a gasoline amount against a beef amount in the column matches, they may have started with a gasoline amount that was approximately matching the beef amount in price. Then they adjusted the gasoline amount downward to account for the lower subjective value of beef. If such adjustment was insufficiently large, they would prefer the gamble with the gasoline match in the check items. In the row matches, the insufficient adjustment away from expected values can explain the bias for risk averse subjects. Risk averse subjects would adjust the certainty equivalent down from the expected value of the gamble. If that adjustment is insufficient,

they would prefer the gamble with the certainty equivalent in the check items. However, insufficient adjustment would predict the opposite effect for risk seeking subjects. Risk seeking subjects would adjust their certainty equivalent not far enough upwards, and consequently tend to prefer the side with the gamble over the side with the certainty equivalent in the check items. Thus anchoring and insufficient adjustment cannot completely explain the bias.

Another possible explanation for the bias is offered by Slovic's (1975) More Important Dimension hypothesis (MID hypothesis). According to this hypothesis, subjects resolve conflicts about matched pairs by preferring the side which is better on the more important attribute. This would explain the bias in the column check items, since the gamble with the matched value had more gasoline, a commodity that was considered the more useful one by all subjects. The bias for the gamble with the certainty equivalent in the row check items may be similarly explained as a tendency to prefer the side that offered more certainty.

So far the experiment does not give any conclusive explanations for the reasons for the strong bias found, but it raises the important question: how well do matches hold up in actual choices when put into different context? Simple matching may involve processes that are different from the judgmental processes involved in the comparison of complex gambles. When confronted with their bias in the

interview after the experiment, several subjects explained it with mismatching. They indicated that they would resolve the inconsistency by changing their match. Other subjects pointed out that they compared the complex gambles element by element, which lead them to different conclusions than the comparisons of total rows and columns required in the matches. They were puzzled by the bias and they were not quickly ready to resolve it either way. If forced, however, they tended to be more suspicious of their comparisons between check item gambles than of their matches.

Naturally, the bias overshadowed some experimental effects. Nevertheless, the more interesting results of the experiment held up against the preference bias. First of all, violations of row and column independence, although neither very large (in the area of 15% dollar difference relative to the higher cash equivalent), were patterned very clearly for all subjects. The first experimental question was: is there any difference in the <u>level</u> of violations of row and column independence? This had to be negated on the basis of the experimental data. But the form of independence violations implied that row independence was systematically violated overriding the bias, while column independence violations were largely due to the bias effect. Particularly interesting is the form of violations of row independence, which suggests that most subjects were multivariate risk averse. The term multivariate risk aversion had been

introduced by Richard (1975) in analogue to the usual (single attribute) definition of risk aversion. It characterizes a behavior that shows preferences among lotteries with equal marginal distributions. Multivariate risk aversion leads to a preference of the gamble that has more balanced outcomes.

The explanation that subjects gave for their multivariate risk aversion is persuasive: they preferred the gamble in which they would win something in both events, over the gamble in which they would win a lot in one event, and nothing in the other. Even when the identical marginal distributions of the check items in the marginality test were pointed out to them, they did not change their preferences. Such multivariate risk aversion could be established as a persistent and intended preference characteristic for all but one subject.

Utility independence as tested in the experiment could not be rejected conclusively. The trend that showed up in the experiment was a peculiar strengthening of preferences of the certainty equivalent, if the context rows were increased. This means that subjects tended to prefer the certain side over the gamble, if they received a bonus in addition to the gamble and the certainty equivalent. One could make an argument that the reverse effect should happen. A bonus in form of some certain commodity amount may decrease the risk aversion of a subject (or increase his risk seeking

attitude), thus shifting his preference to the side of the gamble against which his certainty equivalent was matched. v. Winterfeldt and Fischer (1975) present an example where such a shift in preference appears reasonable. The opposite effect that occurred in the experiment may have been due to the increased perception of a large sure win on the side that included the certainty equivalent.

These types of violations of row and column independence in form of biasses and multivariate risk aversion were very stable and unaffected by instructions or response modes. Preferences, cash equivalents, and ratings showed basically the same response patterns.

The relationships between preference characteristics, such as single attribute risk attitude and substitution rates between commodities and cash, on one hand and violations of independence assumptions on the other, remain less clear. No relationship between (riskless) additivity of commodities--expressed in terms of additivity of the cash equivalents--and row or column independence violations could be observed. Delbeke and Fauville (1974) had hypothesized that row independence in its marginality form would be violated more strongly if attributes were non-additive in the riskless sense. Their experimental data could not support this hypothesis, and neither does this experiment. In fact, the present results suggest that the reason for

violations of marginality (risky additivity) are quite independent from the reasons for violating riskless additivity. Violations of row independence in its marginality form were due to multivariate risk aversion and not to an interaction or complementarity of the commodities.

There are, of course, commodities or attributes for which such relationships may exist. An example are complementary commodities such as shoes. Presumably subjects would prefer a fifty-fifty gamble in which they can win a pair of shoes vs. nothing over a fifty-fifty gamble in which they can win the right vs. the left shoe. Complementarity of commodities leads then to a multivariate risk seeking behavior. In general, multivariate risk attitude will be compounded with riskless non-additivity. But as the experiment indicates, multivariate risk aversion can exist as a pure phenomenon, not influenced by riskless interaction phenomena.

Similar to the relationship between riskless and risky additivity, one can explore the relationships between utility functions derived from riskless trade-offs and utility functions derived from lotteries. In the experiment this comparison could be done by looking at the relationship between substitution rates and risk premia. Both are indicators of the non-linearity of the utility functions. Substitution rates indicate how marginally decreasing the riskless utility for commodities is, risk

premia indicate how convex the risky utility function is. The experimental analysis could identify no relationship between these two indicators. Subjects who had almost linear substitution of gasoline with cash, varied widely in risk attitude about gasoline gambles. Similarly, subjects who all had rather pronounced nonlinear substitution rates of beef with cash were risk neutral, risk seeking, or risk averse.

The results of one pilot subject demonstrate that the concepts of risky and riskless utility function have to be separated quite carefully. This subject stated a certainty equivalent of 3 for a beef gamble (10 0). Since she was risk seeking in the gasoline gamble and also in gambles for cash, this seemed peculiar. As it turned out she had such a strongly decreasing marginal utility for beef, that any increase beyond five pounds was worth nothing to her. The (10 0) beef gamble was worth not much more than a (5 0) beef gamble to her, and her low certainty equivalent said little about her real risk attitude (in the sense of liking or disliking to take chances), but rather about her riskless evaluations of beef. As in the case of the additivity considerations where the risky additivity is compounded with riskless additivity, the risky utility function derived from lottery preferences is compounded with the riskless utility function. This compounding effect may explain that so many subjects had different risk attitudes in the two commodities.

If one accepts the conclusion that there is a compounding relationship between riskless and risky additivity and utility, how would one interpret Tversky's experimental results? First of all, there is no formal or behavioral reason for assuming an identity of riskless or risky utility functions--neither in single attribute shape, nor in aggregation form.

Secondly, the validity of SEU is quite compatible with differences in riskless and risky utility functions. There are some restrictions to these statements however. If the riskless and the risky aggregation rule are additive, then the formal properties of Model 3 imply identical shapes of riskless and risky single attribute utility functions. If they are different, as the utility functions in Tversky's experiment, the SEU model will have to be rejected.

But if, for example, the multiplicative aggregation rule of Model 4 holds, then the SEU assumption implies different shapes of risky and riskless single attribute utility functions. If they are identical, the SEU model would have to be rejected. The experimental results presented here indicate that both aggregation form and shape of riskless and risky utility functions are different. In addition, they show that Model 3 is strongly violated. This still leaves room for Model 4 to be valid which, in fact, is not so strongly violated by the experimental data.

To summarize the discussion of the behavioral interpretation of the experimental results:

1) Both row and column independence are violated through a bias in preferences that contradicts previous matching judgments. This bias may be due either to mismatching or a change in preferences and/or response strategies. This bias questions the validity of Models 1 and 2.

2) For most subjects row independence is systematically violated by multivariate risk aversion. This evidence clearly contradicts the descriptive validity of Model 1 for those subjects. The specific forms of violation of marginality exclude Model 3.

3) Riskless and risky utility functions show independent and different characteristics both in form of aggregation and in the shape of the utility functions in single commodities. These differences are compatible only with Models 2 and 4.

General implications of the experimental results for normative modelling of risky multiattribute preferences are limited by the specific stimuli used in the experiment. Similar phenomena as found in the experimental analysis of preferences among even chance gambles for gasoline and ground beef are likely to occur only in similar stimulus situations, that is when attributes are separable and can be traded off in a simple way with money. But such cases

AD-A048 568 DECISIONS AND DESIGNS INC MCLEAN VA F/G 5/11
EXPERIMENTAL TESTS OF INDEPENCENCE ASSUMPTIONS FOR RISKY MULTIA--ETC(U)
OCT 76 D VON WINTERFELDT N00014-76-C-0074

UNCLASSIFIED NL

2 of 2

ADA048 568

END
DATE
FILMED
2-78
DDC

are not uncommon in real world decision making, and a decision analyst can learn some important lessons from the experimental results for future applications of risky multiattribute utility models in solving such decision problems. The first message is an encouraging repetition of the results of many sensitivity analyses: even in the presence of independence violations, simple models such as additive EV models predict preferences rather well. The experiment showed that these models were normatively wrong in the sense that subjects were single attribute risk averse or risk seeking and that they were multivariate risk averse. Although their expressed preferences among gambles cannot be used as a normative criterion for the validity of the model, one would expect a normative model to covariate with preference judgments in relatively simple tasks. Such covariance was found for most subjects.

Second, the overall level of the observed violations of model assumptions were not very high. But the violations of models--and thus the possible consequences of mismodelling--were larger for some models and/or some stimuli. These differences have to be carefully considered in building and applying multiattribute preference models.

In the discussion of the normative implications of the experimental results, it is important to distinguish between violations of models that represent biasses and violations that are purposeful and consistent. The first type of violations imposes problems of enforcing

consistency, or of finding methods to resolve inconsistencies. The latter type of violations implies questions like: what kind of error does one make when modelling against the expressed intentions of the decision maker's preferences?

An example of the first type of problem is the bias found in the experiment. If subjects actually mismatched, then current procedures for assessing utility functions with such matches have to be considered carefully to avoid such mismatching. One will have to spend much time on the indifference judgment procedures and check them carefully against some more direct consequences of the matches. If, on the other hand, matches were correct but subjects exhibited a bias in the preferences implied by the match, one needs to determine if such biasses are mere inconsistencies or if they themselves contain a purposeful element that should be implemented into the model.

As far as purposeful violations of models are concerned, the violations of marginality by multivariate risk aversion are the most important result of the experiment. In applications it may be useful to determine the degree of multivariate risk aversion to find out how badly the additive Model 3 would be violated.

Since the behavioral phenomena or risky multiattribute preferences were not very well explored, the present experiment began somewhat like a fishing expedition.

Several important characteristics of risky multiattribute preferences could be demonstrated, and their implications for descriptive and normative modelling may be useful. But many questions remain open, and some have been raised by the experiment:

1) What are the precise behavioral relationships between risky and riskless utility functions, their single attribute shapes, and their aggregation forms?
2) What are the reasons behind the exhibited biasses: mismatching, response strategy changes, or real preference changes?
3) What are the behavioral correlates of multivariate risk aversion?

APPENDIX

This appendix lists the left and right cash equivalent responses (CL and CR) for each of the 45 check items, for both sessions, and for each subject. Responses are listed according to the following order of check items:

Check item #	left		right	
1	16	0	0	16
	0	0	0	0
2	16	0	0	16
	0	5	0	5
3	16	0	0	16
	0	10	0	10
4	16	0	0	16
	5	0	5	0
5	16	0	0	16
	5	5	5	5
6	16	0	0	16
	5	10	5	10
7	16	0	0	16
	10	0	10	0
8	16	0	0	16
	10	5	10	5
9	16	0	0	16
	10	10	10	10
10	16	0	x	x
	0	0	0	0
11	16	0	x	x
	0	5	0	5
12	16	0	x	x
	0	10	0	10
13	16	0	x	x
	5	0	5	10
14	16	0	x	x
	5	5	5	5
15	16	0	x	x
	5	10	5	10
16	16	0	x	x
	10	0	10	0
17	16	0	x	x
	10	5	10	5
18	16	0	x	x
	10	10	10	10
19	0	0	x	0
	10	0	0	0
20	0	0	x	0
	10	5	0	5
21	0	0	x	0
	10	10	0	10
22	0	8	x	8
	10	0	0	0
23	0	8	x	8
	10	5	0	5
24	0	8	x	8
	10	10	0	10
25	0	16	x	16
	10	0	10	0
26	0	16	x	16
	10	5	10	5

Check item #	left		right	
27	0	16	x	16
	10	10	0	10
28	0	0	x	0
	10	0	5	0
29	0	0	x	0
	10	5	5	5
30	0	0	x	0
	10	10	5	10
31	0	8	x	8
	10	0	5	10
32	0	8	x	8
	10	5	5	5
33	0	8	x	8
	10	10	5	10
34	0	16	x	16
	10	0	5	0
35	0	16	x	16
	10	5	5	5
36	0	16	x	16
	10	10	5	10
37	0	0	0	0
	10	0	x	x
38	0	8	0	8
	10	0	x	x
39	0	16	0	16
	10	0	x	x
40	8	0	8	0
	10	0	x	x
41	8	8	8	8
	10	0	x	x
42	8	16	8	16
	10	0	x	x
43	16	0	16	0
	10	0	x	x

Check item #	left		right	
44	16	8	16	8
	10	0	x	x
45	16	16	16	16
	10	0	x	x

REFERENCES

Becker, G.M., DeGroot, M.H., and Marschak, J. Measuring utility by a single response sequential method. Behavioral Science, 1964, 226-233.

Becker, G.M. and McClintock, C.M. Value: Behavioral decision theory. Annual Review of Psychology, 1967, 18, 239-286.

Coombs, C.H. A review of the mathematical psychology of risk and risk taking. Michigan Mathematical Psychology Program, MMPP No. 72-6, The University of Michigan, Ann Arbor, 1972.

Dawes, R.M. and Corrigan, B. Linear models in decision making. Psychological Bulletin, 1974, 7, 317-338.

Delbeke, L. and Fauville, J. An empirical test of Fishburn's additivity axiom. Acta Psychologica, 1974, 38, 1-20.

Edwards, W. The theory of decision making. Psychological Bulletin, 1954, 51, 380-417.

Edwards, W. Behavioral decision theory. Annual Review of Psychology, 1961, 12, 473-798.

Farquhar, P.H. Fractional hypercube decompositions of multiattribute utility functions. Technical Report No. 222, Department of Operations Research, Cornell University, Ithaca, New York, 1974.

Fischer, G.W. Four methods for assessing multiattribute utilities: an experimental validation. Technical Report No. 037230-6-T, Engineering Psychology Laboratory, The University of Michigan, Ann Arbor, 1972.

Fischer, G.W. Experimental applications of multiattribute utility models. In: D. Wendt and C. Vlek (eds.): Utility, Probability, and Human Decision Making. Dordrecht, Holland: Reidel, 1975.

Fishburn, P.C. Independence in utility theory with whole product sets. Operations Research, 1965, 13, 28-45.

Fishburn, P.C. Utility theory for decision making. New York: Wiley, 1970.

Fishburn, P.C. Bernoullian utilities for multiple factor situations. In: J.L. Cochrane and M. Zeleney (eds.): Multiple Criteria Decision Making. Columbia, South Carolina: University of South Carolina Press, 1973.

Fishburn, P.C. von Neumann and Morgenstern functions on two attributes. Operations Research, 1974, 22, 35-45.

Fishburn, P.C. and Keeney, R.L. Seven independence concepts and continuous multiattribute utility functions. Journal of Mathematical Psychology, 1974, 11, 294-327.

Fishburn, P.C. Nondecomposable conjoint measurement for bisymmetric structures. Journal of Mathematical Psychology, 1975, 12, 75-89.

Keeney, R.L. Quasi-separable utility functions. Naval Research Logistics Quarterly, 1968, 15, 551-565.

Keeney, R.L. Multiplicative utility functions. Operations Research, 1974, 22, 22-34.

Keeney, R.L. and Raiffa, H. Decision analysis with multiple conflicting objectives: preferences and value trade-offs. New York: Wiley, in press.

Krantz, D.H., Luce, R.D., Suppes, P., and Tversky, A. Foundations of Measurement, Vol. I. New York: Academic Press, 1971.

Lee, W. Decision theory and human behavior. New York: Wiley, 1971.

Luce, R.D. and Krantz, D.H. Conditional expected utility. Econometrica, 1971, 39, 253-271.

von Neumann, J. and Morgenstern, O. Theory of games and economic bahavior. Princeton: Princeton University Press, 1947.

Pollak, R.A. Additive von Neumann and Morgenstern utility functions. Econometrica, 1967, 35, 485-494.

Raiffa, H. Decision analysis. Reading, Mass.: Addison-Wesley, 1968.

Raiffa, H. Preferences for multiattributed alternatives. Rand Memorandum No. RM-5868-DOT-RC. The Rand Corporation, Santa Monica, California, 1969.

Richard, S.F. Multivariate risk aversion, utility independence, and separable utility functions. Management Science, 1975, 22, 12-21.

Savage, L.J. The foundations of statistics. New York: Wiley, 1954.

Slovic, P. Choice between equally-valued alternatives. Journal of Experimental Psychology: Human Perception and Performance, 1975, 1, 280-287.

Tversky. A. Additivity, utility, and subjective probability. Journal of Mathematical Psychology, 1967, 4, 175-202.

Tversky, A. and Kahnemann, D. Judgment under uncertainty: heuristics and biases. In D. Wendt and C. Vleck (eds.): Utility, probability, and human decision making. Dordrecht, Holland: Reidel, 1975.

Wendt, D. Utility and risk. Acta Psychologica, 1970, 34, 214-228.

v. Winterfeldt, D. Multiattribute utility theory: theoretical background and an experimental application. Unpublished diploma thesis, University of Hamburg, Germany, 1971.

v. Winterfeldt, D. and Fischer, G.W. Multiattribute utility theory: models and assessment methods. In D. Wendt and C. Vlek (eds.): Utility, probability, and human decision making. Dordrecht, Holland: Reidel, 1975.

Research Distribution List

Department of Defense

Assistant Director (Environment and Life Sciences)
Office of the Deputy Director of Defense Research and Engineering (Research and Advanced Technology)
Attention: Lt. Col. Henry L. Taylor
The Pentagon, Room 3D129
Washington, DC 20301

Office of the Assistant Secretary of Defense (Intelligence)
Attention: CDR Richard Schlaff
The Pentagon, Room 3E279
Washington, DC 20301

Director, Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, Program Management Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209
(two copies)

Administrator, Defense Documentation Center
Attention: DDC-TC
Cameron Station
Alexandria, VA 22314
(12 copies)

Department of the Navy

Office of the Chief of Naval Operations (OP-987)
Attention: Dr. Robert G. Smith
Washington, DC 20350

Director, Engineering Psychology Programs (Code 455)
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217
(three copies)

Assistant Chief for Technology (Code 200)
Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

Office of Naval Research (Code 230)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Naval Analysis Programs (Code 431)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Operations Research Programs (Code 434)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research (Code 436)
Attention: Dr. Bruce McDonald
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Information Systems Program (Code 437)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research (ONR)
International Programs (Code 1021P)
800 North Quincy Street
Arlington, VA 22217

Director, ONR Branch Office
Attention: Dr. Charles Davis
536 South Clark Street
Chicago, IL 60605

Director, ONR Branch Office
Attention: Dr. J. Lester
495 Summer Street
Boston, MA 02210

Director, ONR Branch Office
Attention: Dr. E. Gloye and Mr. R. Lawson
1030 East Green Street
Pasadena, CA 91106
(two copies)

Dr. M. Bertin
Office of Naval Research
Scientific Liaison Group
American Embassy – Room A-407
APO San Francisco 96503

Director, Naval Research Laboratory
Technical Information Division (Code 2627)
Washington, DC 20375
(six copies)

Director, Naval Research Laboratory (Code 2029)
Washington, DC 20375
(six copies)

Scientific Advisor
Office of the Deputy Chief of Staff
for Research, Development and Studies
Headquarters, U.S. Marine Corps
Arlington Annex, Columbia Pike
Arlington, VA 20380

Headquarters, Naval Material Command (Code 0331)
Attention: Dr. Heber G. Moore
Washington, DC 20360

Headquarters, Naval Material Command (Code 0344)
Attention: Mr. Arnold Rubinstein
Washington, DC 20360

Naval Medical Research and Development Command (Code 44)
Naval Medical Center
Attention: CDR Paul Nelson
Bethesda, MD 20014

Head, Human Factors Division
Naval Electronics Laboratory Center
Attention: Mr. Richard Coburn
San Diego, CA 92152

Dean of Research Administration
Naval Postgraduate School
Monterey, CA 93940

Naval Personnel Research and Development Center
Management Support Department (Code 210)
San Diego, CA 92152

Naval Personnel Research and Development Center (Code 305)
Attention: Dr. Charles Gettys
San Diego, CA 92152

Dr. Fred Muckler
Manned Systems Design, Code 311
Navy Personnel Research and Development Center
San Diego, CA 92152

Human Factors Department (Code N215)
Naval Training Equipment Center
Orlando, FL 32813

Training Analysis and Evaluation Group
Naval Training Equipment Center (Code N-00T)
Attention: Dr. Alfred F. Smode
Orlando, FL 32813

Department of the Army

Technical Director, U.S. Army Institute for the Behavioral and Social Sciences
Attention: Dr. J.E. Uhlaner
1300 Wilson Boulevard
Arlington, VA 22209

Director, Individual Training and Performance Research Laboratory
U.S. Army Institute for the Behavioral and
and Social Sciences
1300 Wilson Boulevard
Arlington, VA 22209

Director, Organization and Systems Research Laboratory
U.S. Army Institute for the Behavioral and Social Sciences
1300 Wilson Boulevard
Arlington, VA 22209

Department of the Air Force

Air Force Office of Scientific Research
Life Sciences Directorate
Building 410, Bolling AFB
Washington, DC 20332

Robert G. Gough, Major, USAF
Associate Professor
Department of Economics, Geography and Management
USAF Academy, CO 80840

Chief, Systems Effectiveness Branch
Human Engineering Division
Attention: Dr. Donald A. Topmiller
Wright-Patterson AFB, OH 45433

Aerospace Medical Division (Code RDH)
Attention: Lt. Col. John Courtright
Brooks AFB, TX 78235

Other Institutions

The Johns Hopkins University
Department of Psychology
Attention: Dr. Alphonse Chapanis
Charles and 34th Streets
Baltimore, MD 21218

Institute for Defense Analyses
Attention: Dr. Jesse Orlansky
400 Army Navy Drive
Arlington, VA 22202

Director, Social Science Research Institute
University of Southern California
Attention: Dr. Ward Edwards
Los Angeles, CA 90007

Perceptronics, Incorporated
Attention: Dr. Amos Freedy
6271 Variel Avenue
Woodland Hills, CA 91364

Director, Human Factors Wing
Defense and Civil Institute of
Environmental Medicine
P.O. Box 2000
Downsville, Toronto
Ontario, Canada

Stanford University
Attention: Dr. R.A. Howard
Stanford, CA 94305

Montgomery College
Department of Psychology
Attention: Dr. Victor Fields
Rockville, MD 20850

General Research Corporation
Attention: Mr. George Pugh
7655 Old Springhouse Road
McLean, VA 22101

Oceanautics, Incorporated
Attention: Dr. W.S. Vaughan
3308 Dodge Park Road
Landover, MD 20785

Director, Applied Psychology Unit
Medical Research Council
Attention: Dr. A.D. Baddeley
15 Chaucer Road
Cambridge, CB 2EF
England

Department of Psychology
Catholic University
Attention: Dr. Bruce M. Ross
Washington, DC 20017

Stanford Research Institute
Decision Analysis Group
Attention: Dr. Allan C. Miller III
Menlo Park, CA 94025

Human Factors Research, Incorporated
Santa Barbara Research Park
Attention: Dr. Robert R. Mackie
6780 Cortona Drive
Goleta, CA 93017

University of Washington
Department of Psychology
Attention: Dr. Lee Roy Beach
Seattle, WA 98195

Eclectech Associates, Incorporated
Post Office Box 179
Attention: Mr. Alan J. Pesch
North Stonington, CT 06359

Hebrew University
Department of Psychology
Attention: Dr. Amos Tversky
Jerusalem, Israel

Dr. T. Owen Jacobs
Post Office Box 3122
Ft. Leavenworth, KS 66027

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 001813-6-T	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Experimental Tests of Independence Assumptions for Multiattribute Preferences		5. TYPE OF REPORT & PERIOD COVERED Technical 6. PERFORMING ORG. REPORT NUMBER SSRI 76-8
7. AUTHOR(s) Detlof von Winterfeldt		8. CONTRACT OR GRANT NUMBER(s) Prime Contract: N00014-76-C-0074 Subcontract: 75-030-0711
9. PERFORMING ORGANIZATION NAME AND ADDRESS Social Science Research Institute University of Southern California Los Angeles, California 90007		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Advance Research Project Agency 1400 Wilson Boulevard Arlington, Virginia 22209		12. REPORT DATE October 1976 13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Office of Naval Research 800 North Quincy Street Arlington, Virginia 22217		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release; Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Support for this research performed by Social Science Research Institute was provided by the Advanced Research Projects Agency of the Department of Defense and was monitored under Contract N00014-76-C-0074 with the Office of Naval Research, under subcontract from Decisions and Designs, Incorporated.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Multiattribute utility
Multivariate risk aversion
Marginality
Risk attitude
Row independence
Column independence
Utility independence

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The purpose of this experiment was to analyze models of human preferences in complex decision situations that are characterized by uncertainty and multiple attributes of outcomes. Four basic models for such risky multiattribute preferences were conisdered, among them the additive and multiplicative expected utility models. Independence assumptions that can test the descriptive validity of these models were formulated.

The validity of the independence assumptions, including the marginality assumption and utility independence, was tested for subjects' preferences

next page

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(continued)

cont.

among even chance gambles for commodity bundles containing gasoline and ground beef. Subjects matched gambles or commodity bundles against a standard and these matches were checked to see if the indifference held in various stimulus contexts as required by the independence assumptions. Effects of response modes, instructions, and personal preference characteristics were examined.

All independence assumptions and models were violated by a bias to prefer a gamble or commodity bundle that was previously matched against a standard, independently of context. Systematic and strong violations of the marginality assumption were found in form of a multivariate risk averion: subjects tended to prefer a gamble with more balanced multiple outcomes over a gamble with extreme multiple outcomes, even if all single outcomes had an equal chance of occurring. Both the bias and multivariate risk aversion were independent of response modes and instructions. Other preference characteristics such as single attribute risk attitude and preferential interaction of commodities seemed unrelated to multivariate risk aversion.

The bias to prefer a previously matched gamble over a standard cannot be explained by any traditional model describing risky multiattribute preferences. This bias could be due either to mismatching or to a change in preferences after matching. The phenomenon of multivariate risk aversion proved to be a stable property of risky multiattribute preferences for the stimuli considered. Descriptive models for risky multiattribute preferences will have to take this phenomenon into account in similar stiumulus situations. For normative modelling the results of the experiment indicate the necessity to carefully check the consistency of preferences assessed by procedures that are based on indifference judgments and to compare them with actual choices. The multivariate risk aversion effect suggests that simple additive expected utility models may, in some cases, be inappropriate for prescribing preferences. Checks of the marginality assumption and analyses of the form of multivariate risk aversion should be designed and tested carefully, before modelling decision makers' preferences with additive expected utility models.